

TITLE Genomic exploration of the hemiascomycetous yeasts: 8:

JOURNAL Zygocaccharomyces rouxii

PUBLISHED FEBS Lett. 487 (1), 52-55 (2000)

REFERENCE 11152883

3 (bases 1 to 1002)

AUTHORS Direct Submission

TITLE Submitted (06-SEP-2000) Genoscope - Centre National de Sequencage,

JOURNAL 2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail: seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)

COMMENT This GSS is part of a random genomic sequencing program of thirteen

yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces

exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,

Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces

lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia

angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,

Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to

5 kb were prepared and both extremities were sequenced. See

keywords for description of this sequence and for the sequence of

the other extremity of this insert.

# FEATURES

source Location/Qualifiers

1..1002 /organism="Zygosaccharomyces rouxii"

/mol\_type="genomic DNA"

/strain="CBS 732"

/db\_xref="taxon:4956"

/clone="AR0A010H04"

/clone\_lib="AR0A"

/notes="end : T3"

misc\_feature <2..5979

/notes="similar to Saccharomyces cerevisiae ORF YCR045c [

similarity to serin proteases ]"

/evidence=not\_experimental

# ORIGIN

Alignment Scores: 1.88e-09 Length: 1002

Pred. No.: 198.00 Matches: 79

Score: 43.57% Conservative: 43

Best Local Similarity: 28.21% Mismatches: 98

Query Match: 8.80% Indels: 60

DB: 29 Gaps: 15

US-09-985-689a-6 (1-434) x CNS06D68 (1-1002)

QY 14 GlnSerSerTyrGlyLeuTyr-----GlyGlnGlyGlnValValAla 27

Db 278 GAAGAGAGTTTCAATTATTACTATTATAATGCGACAGGGTAGAATAATCAATGCATAT 337

QY 28 ValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisGluAlaPheArg 47

Db 338 ATATTGATATCTGGAATC-----TATAAGAACATAGGACTTCTGT 379

QY 48 GlyLysIleThrAlaIleTyrAlaLeuGlyArgThrAsnAsnAla---AsnAspProAsn 66

Db 380 GGACGT-----GCCATATTCGCCCGAGACTTTACAGGGAGAGGTCTCGTGCATCTAT 433

QY 67 GlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGlyThrSerAsnLysGlyMet 86

Db 434 GGCATGGAACACTCACGTGTGTCGATGTG-----GGATCGAGCAACTTGTGT 484

QY 87 AlaProGlnAlaAsnLeuValPheGlnSerValMetAspSerAsnGlyGlyLeuGlyGly 106

Db 485 GCTAAAAGGTGAATTAATTAGTTAGTTAAGCTTTG---NACATAGAGCCCAAGCAAT 541

QY 107 LeuProSerAsnValSerThrLeu-----PheSerGlnAlaTyrSerAlaGlyAla 123

Db 542 TTAACAACAGTAATTAGCGTGTGTGAATTTGCCGTGAATCATTTGTAATCATCGGCAAA 601

QY 124 Arg-----IleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyrThrAsp 141

Db 602 AAGGTTGCGTTGCCAATCTTTTCATGGTGGCGCTGAGAAACAGTGTATAATCAAGCA 661

QY 142 SerArgAsnValAspAspTyrValArgLysAsnAspMetAlaValLeuPheAlaAlaGly 161

Db 662 ATTAAA-----GCTGCTCATGACAGCGCTCATATTATTAGTTGCGACAGGT 709

QY 162 AsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThr 181

Db 710 AATTCTAATATAT--GCATGTGGNAATAGCCTCGCTCTGCACCTGAGGCTATTACT 766

QY 182 ValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsnHis 201

Db 767 GTAGGCGCC-----TTTGATGATCGGTACACATACA 796

QY 202 ValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspValMet 221

Db 797 ATTGCATAATTAGTAATTCGGGTCATGT-----GTTGATATTTT 838

QY 222 AlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrp 241

Db 839 GCATCAGCGCTRAAAGTTAAATCATTG--TCTGCATTCACACCT-----880

QY 242 AlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIleVal 261

Db 881 -----CATAAACCAATTGCATTTTCAGGTACTTCGATGGCTTCACCAAGTGT 928

QY 262 AlaGlyAsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrProLys 281

Db 929 ACTGGACTTGTTGCT-----ATATTATTGGATAAAGCGTTGAACACAGAA 973

RESULT 11

TA315H10P 574 bp DNA linear GSS 13-DEC-2000

LOCUS T. brucei sheared genomic DNA clone 315h10, forward sequence,

DEFINITION genomic survey sequence.

ACCESSION AL490202

VERSION AL490202.1 GI:11866292

KEYWORDS GSS.

SOURCE Trypanosoma brucei

ORGANISM Trypanosoma brucei

REFERENCE 1 (bases 1 to 574)

AUTHORS Hall, N., Bowman, S., Lennard, N.J., Doggett, J., Atkin, R.,

Chillingworth, C., Ormond, D., Harris, B., El-Sayed, N., Hou, L.,

Melville, S.E., Rajandream, M.A. and Barrell, B.G.

TITLE Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing

JOURNAL project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton,

COMMENT Cambridge CB10 1SA, E-mail: barrellesanger.ac.uk and

nh@sanger.ac.uk

Constructed at the Institute for Genomic Research (TIGR),

Rockville, MD. Genomic DNA isolated from a cloned population of

Trypanosoma brucei (TREU927/4 Gutat 10.1) was mechanically sheared

to give a tight size distribution (

4 kb). The v + i method used for the library construction is

described in detail in Smith, H. and Venter, J.C. (Making small

insert libraries for whole genome shotgun sequencing projects. In

Genome Sequencing: A Practical Approach, eds. M. Vaudin and B.

Barrell, Oxford University Press, 1999).

Email: nelsayed@tigr.org

Details of T. brucei sequencing at the Sanger Centre are available

at http://www.sanger.ac.uk/Projects/T\_brucei/.

FEATURES Location/Qualifiers

1..574

/organism="Trypanosoma brucei"

/mol\_type="genomic DNA"

/strain="TREU927"

/db\_xref="taxon:5691"

/clone="315h10"

ORIGIN

Alignment Scores: 4.27e-09 Length: 574

Pred. No.:



AUTHORS
Harris,L.J., Glasco,T., Rocheleau,H., Allard,S., Chapados,J.,
Couroux,P., De Moors,A., Hattori,J.I., Ouellet,T., Robert,L.S.,
Singh,J.A., Sproull,D. and Linker,N.A.
Expressed Sequence Tags from Fusarium graminearum mycelium
Unpublished (2001)
Contact: Harris, Linda J.
Eastern Cereal and Oilseed Research Centre
Agriculture and Agri-food Canada
Bldg 21, Central Experimental Farm, Ottawa, Ontario, K1A 0C6,
CANADA
Tel: (613) 759-1314
Fax: (613) 759-6566
Email: harris1j@em.agr.ca.
FEATURES
Location/Qualifiers
1..718
/organism="Gibberella zeae"
/mol\_type="mRNA"
/strain="DAOM 180378"
/db\_xref="taxon:5518"
/clone="Fg02.10G08"
/tissue\_type="Mycelial tissue"
/dev\_stage="Asexual"
/lab\_host="E. coli (Sure cells)"
/clone\_lib="Fg02.AAPC.ECORC.Fusarium\_graminearum\_mycelium"
/notes="Vector: Bluescript SK+/XhoI-EcoRI; Site 1: EcoRI;
Site 2: XhoI; Mycelial tissue was collected from v8 agar
plates after a growth period of 6-7 days at 25 C with 14
hrs (FL/UV) day light exposure. Mycelia was ground in
liquid nitrogen prior it's storage at -80 C until RNA
extraction. Directional cloning with 5' end of cDNA cloned
into EcoRI site of pBluescript and 3' end of cDNA cloned
into XhoI site of pBluescript (Stratagene, La Jolla, CA)."
ORIGIN
Alignment Scores:
Pred. No.: 8.98e-09 Length: 718
Score: 189.00 Matches: 66
Percent Similarity: 43.05% Conservative: 30
Best Local Similarity: 29.60% Mismatches: 89
Query Match: 8.40% Indels: 38
DB: 12 Gaps: 10
US-09-985-689A-6 (1-434) x BI750157 (1-718)
QY 46 PheargglylslethrAlaIleTyrAlaLeuGlyArgThrAsnAsnAlaAsnAspPro 65
Db 9 TTCAGAGSTGCTGCTCAGGCTGTCTACATGCTTACGGCGCCAGACGCT---GACACC 65
QY 66 AsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn-----GlyThrSerAsn 83
Db 66 AACGGTCACGGAACCTACGTTGCTGGCATTATTCGCGAAGACATACGCTGTGCCAAG 125
QY 84 LysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerValMetAspSerAsnGly--- 102
Db 126 AAGGCCACCATC---CAAGCTGTCAAGTCTTCCAGGCTAGTTCATCCAGCCTCCATC 182
QY 103 GlyLeuGlyGlyLeuProSerAsnValSerThrLeuPheSerGlnAlaTyrSerAlaGly 122
Db 183 ATCTCTCGTTCACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 227
QY 123 AlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyrThrThrAspSer 142
Db 228 GCGCGAACCAGACATCAGTCAGTCGTCAATATGTCTCTCGCGCTGCTTACTCTGCTCTCTC 287
QY 143 ArgAsnValAspAspTyrValArgLeuAsnAspMetAlaValLeuPheAlaAlaGlyAsn 162
Db 288 AACACGCTGTCAGTCTKCTCAGCTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 347
QY 163 GluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaValAsnAlaIleThrVal 182
Db 348 GATGGTGCCACGCTGCGCAACACTCT---CCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 404
QY 183 GlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsnHisVal 202

Db 405 GGTGCCATTGACAGC-----AATGGGCC-----ATT 431
QY 203 AlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspValMetAla 222
Db 432 KCCTCGTACTCCAACTACGCTACGCTGCTC-----GATACTTTGCT 473
QY 223 ProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrpAla 242
Db 474 CCTGCACACGCGTCTCTCCGCC-----TGCTAC 503
QY 243 AsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIleValAla 262
Db 504 ACCAGCAACAGTCCCAACACCATCAGCGGACACCTCCATGCTGCTCTCCATTCGCTGCC 563
QY 263 GlyAsnVal 265
Db 564 GGAATTGTC 572

RESULT 14
BZ893395/c
LOCUS
DEFINITION
BZ893395 716 bp DNA linear GSS 30-JUL-2003
genomic survey sequence.
ACCESSION
BZ893395
VERSION
BZ893395.1 GI:33343985
KEYWORDS
GSS.
SOURCE
Halorubrum lacusprofundi
ORGANISM
Halorubrum lacusprofundi
Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
Halobacteriaceae; Halorubrum.
REFERENCE
1 (bases 1 to 716)
AUTHORS
Gao,Y., Roach,J., Glusman,G., Baliga,N.S., Deutsch,K., Pan,M.,
DasSarma,S., Ng,W.V. and Hood,L.
Low-pass Sequencing for Microbial Comparative Genomics
Unpublished (2003)
CONTACT: Gao Y
Institute for Systems Biology
1441 North 34th Street, Seattle, WA 98103, USA
Tel: 206 732 1412
Fax: 206 732 1299
Email: ygao@systemsbiology.org
Seq primer: M13 Forward
Class: shotgun.
FEATURES
Location/Qualifiers
1..716
/organism="Halorubrum lacusprofundi"
/mol\_type="genomic DNA"
/strain="ATCC 49239"
/db\_xref="taxon:2247"
/clone\_lib="H1 pUC18 Library"
/note="Vector: pUC18; Site 1: SmaI; A shotgun library was
constructed from Halorubrum lacusprofundi genomic DNA
using pUC18/SmaI/BAP plasmid"
ORIGIN
Alignment Scores:
Pred. No.: 1.82e-08 Length: 716
Score: 186.00 Matches: 72
Percent Similarity: 48.18% Conservative: 34
Best Local Similarity: 32.73% Mismatches: 77
Query Match: 8.27% Indels: 38
DB: 28 Gaps: 10
US-09-985-689A-6 (1-434) x BZ893395 (1-716)
QY 60 AsnAsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn 79
Db 701 TCCGATGCTCGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 642
QY 80 GlyThrSerAsnLys-----GlyMetAlaProGlnAlaAsnLeuValPheGlnSerVal 97
Db 641 AACGGAGTGGGACCGCAATCGCTCGACGACGACGACGACGACGACGACGACGACGACG 582





Qy 275 AsnArgGlyIleThrProLysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAla 293  
Db 584 CGGAGCTGGTG-----AATCCTGCCAGTGTGAAGCACCGCTTGTATAGCGTCAGCC 634

Search completed: April 4, 2004, 11:55:12  
Job time : 2297.84 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: April 4, 2004, 08:10:21 ; Search time 327.248 Seconds  
(without alignments)  
4948.852 Million cell updates/sec

Title: US-09-985-689A-7  
Perfect score: 2252  
Sequence: 1 NDVARGIVKADVAQNNFGLY.....EVOAYNPVSPQTPSLAIHV 433

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2466186 seqs, 1870095128 residues  
Total number of hits satisfying chosen parameters: 4932372

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ p2n.model -DBV=xlh  
-Q=/cgn2\_1/USPTO.spool/US0985689/runat\_31032004\_161809\_4271/app.query.fasta\_1.3498  
-DB=Published Applications NA -QFT=fastap -SUFFIX=rnpb -MINMATCH=0.1  
-LOOPEXT=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62  
-TRANS=human40.cdi -LIST=45 -DOCLIGN=200 -THR SCORE=ptc -THR MAX=100  
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0  
-MAXLEN=2000000000 -USER=US0985689@cgn\_1\_1\_601@runat\_31032004\_161809\_4271  
-NCPU=6 -ICPU=3 -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBIOCK=100  
-LONGLOG -DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA:  
1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:  
2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq:  
3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq:  
4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq:  
5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq:  
6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq:  
7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq:  
8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:  
9: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq:  
10: /cgn2\_6/ptodata/1/pubpna/US09B\_PUBCOMB.seq:  
11: /cgn2\_6/ptodata/1/pubpna/US09C\_PUBCOMB.seq:  
12: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq:  
13: /cgn2\_6/ptodata/1/pubpna/US10A\_PUBCOMB.seq:  
14: /cgn2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq:  
15: /cgn2\_6/ptodata/1/pubpna/US10C\_PUBCOMB.seq:  
16: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq:  
17: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq:  
18: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq:  
19: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Query Match	Length	DB ID	Description
-----				

1	2125.5	94.4	1305	15	US-10-385-662-1	Sequence 1, Appli
2	432.5	19.2	1977	13	US-10-090-624-11	Sequence 11, Appl
3	396	17.6	1236	13	US-10-090-624-2	Sequence 2, Appli
4	396	17.6	1962	13	US-10-090-624-15	Sequence 15, Appl
5	355	15.8	3624	14	US-10-156-761-5701	Sequence 5701, Ap
6	355	15.8	9025608	14	US-10-156-761-1	Sequence 1, Appli
7	332	14.7	3417	14	US-10-156-761-3306	Sequence 3306, Ap
8	332	14.7	9025608	14	US-10-156-761-1	Sequence 1334, Ap
9	325	14.4	1329	9	US-09-974-300-1934	Sequence 1, Appli
10	323	14.3	135638	14	US-10-314-657-1	Sequence 5, Appli
11	311.5	13.8	4765	13	US-10-090-624-5	Sequence 113, App
12	297	13.2	1560	15	US-10-084-846A-113	Sequence 1, Appli
13	297	13.2	59816	15	US-10-084-846A-2	Sequence 2, Appli
14	297	13.2	59816	15	US-10-084-846A-1	Sequence 29, Appl
15	279.5	12.4	3743	10	US-09-927-827-23	Sequence 33, Appl
16	276.5	12.3	3788	10	US-09-927-827-33	Sequence 1, Appli
17	271.5	12.1	1306	9	US-09-966-921A-1	Sequence 5, Appli
18	271.5	12.1	1330	14	US-09-966-921A-5	Sequence 5384, Ap
19	259.5	11.5	3303	14	US-10-156-761-5384	Sequence 17, Appl
20	258	11.5	2166	12	US-10-344-231-17	Sequence 17, Appl
21	258	11.5	2166	12	US-10-363-332A-17	Sequence 1335, Ap
22	254.5	11.3	1971	9	US-09-974-300-1935	Sequence 112429,
23	254.5	11.3	2192	12	US-10-424-599-112429	Sequence 1338, Ap
24	253.5	11.3	1485	9	US-09-974-300-1938	Sequence 34, Appl
25	252	11.2	3884	10	US-09-927-827-34	Sequence 1, Appli
26	245.5	10.9	840	14	US-10-209-812-1	Sequence 11, Appl
27	245.5	10.9	1140	8	US-08-322-678-11	Sequence 12, Appl
28	245.5	10.9	1140	16	US-10-323-324-11	Sequence 12, Appl
29	245.5	10.9	1140	16	US-10-323-324-12	Sequence 6, Appli
30	245.5	10.9	1140	16	US-10-313-853-6	Sequence 20, Appl
31	245.5	10.9	1143	14	US-10-344-231-20	Sequence 20, Appl
32	245.5	10.9	2588	12	US-10-363-332A-20	Sequence 30, Appl
33	245.5	10.9	2588	12	US-09-927-827-30	Sequence 18, Appl
34	236.5	10.5	3452	10	US-10-344-231-18	Sequence 3, Appli
35	232.5	10.3	2267	12	US-10-363-332A-18	Sequence 953, App
36	232.5	10.3	2267	12	US-09-891-711-3	Sequence 5689, Ap
37	232	10.3	4338	14	US-10-156-761-5689	Sequence 5, Appli
38	232	10.3	4338	12	US-10-342-887-953	Sequence 9, Appli
39	230	10.2	1332	14	US-10-156-761-5689	Sequence 6, Appli
40	228	10.1	4198	9	US-09-891-711-5	Sequence 5, Appli
41	226	10.0	1140	15	US-10-146-905A-9	Sequence 9, Appli
42	224	9.9	1497	8	US-08-322-678-6	Sequence 6, Appli
43	224	9.9	1497	9	US-09-060-854B-1	Sequence 1, Appli
44	224	9.9	1497	14	US-10-033-325-1	Sequence 1, Appli
45	224	9.9	1497	14	US-10-228-572-1	Sequence 1, Appli

#### ALIGNMENTS

RESULT 1  
US-10-385-662-1  
; Sequence 1, Application US/10385662  
; Publication No. US2004002432A1  
; GENERAL INFORMATION:  
; APPLICANT: OKUDA, MITSUYOSHI  
; APPLICANT: SAITO, TSUYOSHI  
; APPLICANT: SAITO, KAZUHIRO  
; APPLICANT: SUMITOMO, NOBUYUKI  
; APPLICANT: IZAWA, YOSHIFUMI  
; APPLICANT: SAEKI, KATSUFUMI  
; APPLICANT: KOBAYASHI, TOHRU  
; APPLICANT: NOMURA, MASATUMI  
; TITLE OF INVENTION: Alkaline protease  
; FILE REFERENCE: 234938USO  
; CURRENT APPLICATION NUMBER: US/10/385.662  
; CURRENT FILING DATE: 2003-03-12  
; PRIOR APPLICATION NUMBER: JP 2002-081428  
; PRIOR FILING DATE: 2002-03-22  
; PRIOR APPLICATION NUMBER: JP 2002-165987  
; PRIOR FILING DATE: 2002-06-06  
; PRIOR APPLICATION NUMBER: JP 2002-304230  
; PRIOR FILING DATE: 2002-10-18  
; PRIOR APPLICATION NUMBER: JP 2002-304231

;; PRIOR FILING DATE: 2002-10-18  
;; NUMBER OF SEQ ID NOS: 20  
;; SOFTWARE: PatentIn version 3.1  
;; SEQ ID NO 1  
;; LENGTH: 1305  
;; TYPE: DNA  
;; ORGANISM: Bacillus sp. KSM-KP43  
;; FEATURE:  
;; NAME/KEY: CDS  
;; LOCATION: (1)..(1305)  
;; OTHER INFORMATION:  
US-10-385-662-1

Alignment Scores:  
Pred. No.: 8,02e-222 1305  
Score: 2125.50  
Percent Similarity: 97.93%  
Best Local Similarity: 93.55%  
Query Match: 94.38%  
DB: 15  
Gaps: 1  
Indels: 1  
Mismatch: 8  
Conservative: 19  
Matches: 406  
Length: 1305

US-09-985-689a-7 (1-433) x US-10-385-662-1 (1-1305)

Qy 1 AsnAspValAlaAArgGlyIleValLysAlaAAspValAlaGlnAsnAsnPheGlyLeuTyr 20  
Db 1 AATGATGTGGCGTGGAAATGTGCAAGCGATGTGGCTCAGAGCAGCTACGGTGTAT 60  
Qy 21 GlyGlnGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40  
Db 61 GGACAAGGACAGATCGTAGCGGTGGCCGATACAGGGCTTGATACAGTCGCAATGACAGT 120  
Qy 41 SerMetHisGluAlaPheAArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60  
Db 121 TCGATGATGAAGCTTCGCGGGAATTAATCTGCAATATATGATTTGGCAGCGAAT 180  
Qy 61 AsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn--- 79  
Db 181 AATGCCAATGATACGAATGGTTCATGGTACGATGCTGGCTGGCTCCGTTATAGAAACGGC 240  
Qy 80 AlaThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 99  
Db 241 TCCACTAATAAGAAATGGCCCTCAGCGAATCTAGTCTCCATCTCATCTATGATAGC 300  
Qy 100 GlyGlyGlyLeuGlyGlyLeuProAlaAsnLeuGlnThrLeuPheSerGlnAlaTyrSer 119  
Db 301 GGTGGGGACTTGGAGGACTACCTTCGAATTCGAAACCTTATTTCAGCCCAAGCATACAGT 360  
Qy 120 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyrThr 139  
Db 361 GTGTGTCCGAATTCATCAAACTCTCGGAGCAGCAGTGAATGGGCTTACACAACA 420  
Qy 140 AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 159  
Db 421 GATTCAGAAATGTGGATGACTATGTGGCAAAATGATATGACGATCCTTTTCGTGCC 480  
Qy 160 GlyAsnGlyGlyProGlySerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 179  
Db 481 GGAATGAAGGACCGAAACGGCGAACCATCATGTCAGCCAGCAGCTAAATAATGCAATA 540  
Qy 180 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 199  
Db 541 ACAGTCGGAGCTACGGAAACCTCGCCCAAGCTTTGGTCTTATCGGACAATATCAAC 600  
Qy 200 HisValAlaGlnPheSerSerArgGlyProThrArgAspGlyArgIleLysProAspVal 219  
Db 601 CATGTGGCACAGTCTCTTCAGTCGACCGAAGGATGACGAGTCAAAACCGGATGTC 660  
Qy 220 MetAlaProGlyThrTyrIleLeuSerAlaArgSerSerLeuAlaProAspSerPhe 239  
Db 661 ATGGCACCGGGAACGTTTATATATCATCATCATCTCTCTGACCGGATCTCTCTCTC 720  
Qy 240 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProIle 259

Db 721 TGGCGAACCATGACAGTAATAATATGATATGATGGTGGAGCTCCATGGCTACACCATC 780  
Qy 260 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyValThrPro 279  
Db 781 GTTCTGAAACAGTGGCACAGCTTCGTGAGCATTTTCTGAAAAACAGAGCATCACACCA 840  
Qy 280 LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspValGlyLeuGlyPhe 299  
Db 841 AAGCCTTCTCTATTAAAGCGGCACTGATTTGCCGTGCGAGCTGACATCGGCTTGGCTAC 900  
Qy 300 ProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaPhe 319  
Db 901 CCGAACGGTAACCAAGATGGGACAGTGCATTTGATGATAATCCCTGAACGTTGCCAT 960  
Qy 320 ValAsnGluThrSerProLeuSerThrSerGlnLysAlaThrTyrSerPheThrAlaGln 339  
Db 961 GTGAACGAGTCCAGTCTCTTATCCACCAGCCAAAGACGACTACTCTGTTTACTGCTACT 1020  
Qy 340 AlaGlyLysProLeuLysIleSerLeuValTrpSerAspAlaProGlySerThrThrAla 359  
Db 1021 GCCGCAAGCCTTTGAAATCTCCTGATGCTGATGCTGCTGCTGCTGCTGCTGCTGCT 1080  
Qy 360 SerLeuThrLeuValAsnAspLeuValIleThrAlaProAsnGlyThrLysTyr 379  
Db 1081 TCCGTAAACGCTTGCAATGATCTGGACCTTGCTATTACCGCTCCAAATGGCACAGTAT 1140  
Qy 380 ValGlyAsnAspPheThrAlaProTyrAspAsnAsnTrpAspGlyArgAsnAsnValGlu 399  
Db 1141 GTAGAAATGACTTACTTCCCATACATGATGATGATGATGATGATGATGATGATGAT 1200  
Qy 400 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrValGluValGlnAlaTyrAsn 419  
Db 1201 AATGATTTATTAAATGACCAAAAGCGGACGATATCAATTTAGGTACAGGCTTATAAC 1260  
Qy 420 ValProValSerProGlnThrPheSerLeuAlaIleValHis 433  
Db 1261 GTACCGTGGACCAACAGACCTTCTCTGTCGAATTTGTAAT 1302

## RESULT 2

US-10-090-624-11  
; Sequence 11, Application US/10090624  
; Publication No. US20020132335A1  
; GENERAL INFORMATION:  
; APPLICANT: TAKAKURA, Hikaru  
; APPLICANT: MORISHITA, Mio  
; APPLICANT: SHIMOUO, Tomoko  
; APPLICANT: ASADA, Kiyozo  
; APPLICANT: KATO, Ikunoshin  
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE  
; FILE REFERENCE: TAKAKURA=6  
; CURRENT APPLICATION NUMBER: US/10/090,624  
; PRIOR FILING DATE: 2002-03-06  
; PRIOR APPLICATION NUMBER: 09/445,472  
; PRIOR FILING DATE: 1999-12-06  
; PRIOR APPLICATION NUMBER: 151969/1997  
; PRIOR FILING DATE: 1997-06-10  
; NUMBER OF SEQ ID NOS: 33  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 11  
; LENGTH: 1977  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Synthetic  
US-10-090-624-11

Alignment Scores:  
Pred. No.: 6.62e-37 1977  
Score: 432.50  
Percent Similarity: 44.85%  
Best Local Similarity: 30.66%  
Query Match: 19.21%  
DB: 13  
Indels: 85  
Gaps: 16  
Matches: 134  
Conservative: 62  
Mismatch: 156

US-09-985-689A-7 (1-433) x US-10-090-624-11 (1-1977)

QY 8 ValLysAlaAspValAlaGlnAsnAsnProGlyLeuTyrGlyGlnIleValAla 27  
DB 433 ATAGGGCGGATACCGCTCGGACCTCCCTCGGCTACGACGAGCGGTGTGGTGGTCC 492

QY 28 ValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerMetHisGluAlaPheArg 47  
DB 493 ATCGTCGATACGGGTATACGCGAAC-----CACCCCGATCTGAAG 534

QY 48 GlyLysIleThrAlaLeuTyr---AlaLeuGlyArgThrAsnAsnAlaAsnAspProAsn 66  
DB 535 GGCAGGTCATAGGTCGTGACGCGCGTCAACGCGAGGTGCGACCCCTACGATACGACG 594

QY 67 GlyHisGlyThrHisValAlaGlySerValLeuGly-----AsnAlaThrAsn 82  
DB 595 GGCACGGAACCCACGTCGCGGTATCGTTCGCGAACCGCGGACGCTTAACCTCCAGTAC 654

QY 83 LysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMet-----AspSerGly 100  
DB 655 ATAGGCGTCGCCCGCGCGAGGCTCGTCGCGGTCAAGGTTCCTCGGTGCGACGCTTCG 714

QY 101 GlyGlyLeuGlyLeuProAlaAsnLeuGlnThrLeuPheSerGlnAlaTyrSerAla 120  
DB 715 GGAAGCGTCCACCATATCGCGGTGTGACTGGTCTGCTCAGAACGAGCAAGTAC 774

QY 121 GlyAlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyrThrAsp 140  
DB 775 GGGATAAGGTCATCAACCTCTCCCTCGGCTCTCCAGAGCTCCGACGCGAACCGACTCC 834

QY 141 SerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAlaGly 160  
DB 835 CTCAGTCAGCGCTCAACACCGCTGGAGCGCGGTATAGTAGTCTGCTGCGCCCGCGC 894

QY 161 AsnGluGlyProGlySerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThr 180  
DB 895 AACAGCGCGCGAACACCTACCGCTCGGCTCACCGCGCGCGAGCAAGTCAATACC 954

QY 181 ValGlyAlaThrGlnAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsnHis 200  
DB 955 GTCGGTGC-----GTTGACGACCAACGACCAAC 981

QY 201 ValAlaGlnPheSerSerArgGlyProThrArgAspGlyValGlyLeuAspValMet 220  
DB 982 ATCGCCAGCTTCTCCAGCGGGGACCGACCGCGGAGGCTCAAGCGGAGTCTGTC 1041

QY 221 AlaProGlyThrTyrIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrp 240  
DB 1042 GCGCCGCGGCTTGACATCATAGCCCGCGCGCGCAGC-----GGAACCAAGCATGGGC 1092

QY 241 AlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIleVal 260  
DB 1093 ACCCCGATAAACGACTACTACCAAGGCTCTGGNACCAAGCATGGCCACCCCGACGTT 1152

QY 261 AlaGly---AsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyValThrPro 279  
DB 1153 TCGGGGCTTGGCGCTCATCTCTCCAGGCGCC-----CGG 1188

QY 280 Lys-----ProSerLeuLeuLysAlaAlaLeuIleAlaGlyAla----- 292  
DB 1189 AGCTGGACCCCGGACCAAGGTGAAGACCGCCCTCATCGACCGCGGCACATAGTCGCCGCC 1248

QY 293 -----AlaAspValGlyLeuGlyPheProAsnGlyAsnGlnGlyTrpGlyArgVal 309  
DB 1249 AAGGAGATACGCGACATCGCTACGTTGGC-----GGTAGGTG 1287

QY 310 ThrLeuAspLysSerLeu-----AsnValAlaPheValAsnGluThrSerProLeu 326  
DB 1288 AACGTCCTACAGCCCATAGTACGACGAGTACGCAAGCTCACCTTCACCGGCTCGGTC 1347

QY 327 SerThrSerGlnLysAlaThrTyrSerPheThrAlaGlnAlaGlyLysProLeuLysIle 346  
DB 327 SerThrSerGlnLysAlaThrTyrSerPheThrAlaGlnAlaGlyLysProLeuLysIle 346

DB 1348 GCGCACAGGGAAGCGCCACCCACACCTTC-----1377

QY 347 SerLeuValTrpSerAspAlaProGlySerThrThrAlaSerLeuThrLeuVal-----364

DB 1378 -----GACGTCAGCGCGCCACCTTCGTGACCCGCCCTCTACTGGGAC 1422

QY 365 -----AsnAspLeuAspLeuValIleThrAlaProAsnGlyThrLysTyrValGly 381  
DB 1423 ACGGCTCTGAGCGACATCGACCTCTACTCTACGACCCCAACGGAACGAG---GTTGAC 1479

QY 382 AsnAspPheThrAlaProTyrAspAsnTrpAspGlyArgAsnAsnValGluAsnVal 401  
DB 1480 TACTCTACACCGCTCTACTAC-----GGCTTCGAGAAGGTC 1515

QY 402 PheIleAsnAlaProGlnSerGlyThrTyrThrValGluValGlnAlaTyr 418  
DB 1516 GCCTACTACAACCGACCGCGCGAACCTGACGCTCAAGGTCGTGAGTAC 1566

RESULT 3

US-10-090-624-2  
; Sequence 2, Application US/10090624  
; Publication No. US20020132335A1  
; GENERAL INFORMATION:  
; APPLICANT: TAKAKURA, Hikaru  
; APPLICANT: MORISHITA, Mio  
; APPLICANT: SHIMOJO, Tomoko  
; APPLICANT: ASADA, Kiyozo  
; APPLICANT: KATO, Ikunoshin  
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE  
; FILE REFERENCE: TAKAKURA=6  
; CURRENT APPLICATION NUMBER: US/10/090,624  
; PRIOR FILING DATE: 2002-03-06  
; PRIOR APPLICATION NUMBER: 09/445,472  
; PRIOR FILING DATE: 1999-12-06  
; PRIOR APPLICATION NUMBER: 151969/1997  
; PRIOR FILING DATE: 1997-06-10  
; NUMBER OF SEQ ID NOS: 33  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 2  
; LENGTH: 1236  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Synthetic  
US-10-090-624-2

Alignment Scores:  
Pred. No.: 3,22e-33 Length: 1236  
Score: 396.00 Matches: 127  
Percent Similarity: 42.56% Conservative: 56  
Best Local Similarity: 29.53% Mismatches: 163  
Query Match: 17.58% Indels: 84  
DB: 13 Gaps: 15

US-09-985-689A-7 (1-433) x US-10-090-624-2 (1-1236)

QY 16 AsnPheGlyLeuTyrGlyGlnIleValAlaValAlaAspThrGlyLeuAspThr 35  
DB 58 AACTTGGGATATGATGTTCTTGGAAATCAATAAGTAATATGACACTGGAATTGAC---114

QY 36 GlyArgAsnAspSerSerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAla 55  
DB 115 -----GCTTCTCATCCAGATCTCCAGGAAGAAGTA-----144

QY 56 LeuGlyArgThrAsnAsnAlaAsn-----AspProAsnGlyHisGly 69  
DB 145 ATTGGGTGGGTAGATTTTGTCAATGGTAGGAGTTATCCATACGATGACCATGGACATGGA 204

QY 70 ThrHisValAlaGlySerValLeuGly-----AsnAlaThrAsnLysGly 84  
DB 205 ACTCATGTAGTCTCAATAGCAGCTGGTACTGGAGCAGCAAGTAATGCAAGTACCAAGGA 264

QY 85 MetAlaProGlnAlaAsnLeuValPheGlnSerIleMet-----AspSerGlyGly 102

```
Db 265 ATGGCTCCAGGAGTAAGCTGGCGGAATTAAGGTTCTAGTCCGATGTTCTGGAAGC 324
Qy 103 LeuGlyGlyLeuProAlaAsnLeuGlnThrLeuPheSerGlnAlaTyrSerAlaGlyAla 122
Db 325 ATATCTACTATATAATTAAGGAGTGTAGTGGCGCGTGTGATAACAAGATAAGTACGGAAT 384
Qy 123 ArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyrThrThrAspSerArg 142
Db 385 AGGTCAATTAATCTTTCTTTGGTTCAAGCCAGCTCAGATGTTGACGCTCTAAGT 444
Qy 143 AsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAlaGlyAsnGlu 162
Db 445 CAGGCTGTTAATGTCAGCGTGGATGCTGATTAAGTGTGCTGGCGTGGAAACAGT 504
Qy 163 GlyProGlySerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrValGly 182
Db 505 GGACCTAACAGTATACAACTGGTCTCCAGCGAGCTGCAAGCAAGTTATTACAGTTGA 564
Qy 183 AlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsnHisValAla 202
Db 565 GCC-----GTTGCAAGATGATGATTATAACA 591
Qy 203 GlnPheSerSerArgGlyProThrArgAspGlyArgIleLysProAspValMetAlaPro 222
Db 592 AGCTTCTCAAGCAGAGGCCCACTGCAGCGGCGCTTAAGCGTGAAGTGTGTGCTCCA 651
Qy 223 GlyThrTyrIleLeuSerAlaArgSerSerLeuAlaProAspSerPheTrpAlaAsn 242
Db 652 GGAACCTGATTAATGCTGCGCAGAGCAAGT-----GGAACCTAGCAGTGGTCAACA 702
Qy 243 HisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIleValAlaGly 262
Db 703 ATTAATGACTATTACACAGCAGCTCCTGGGACATCAATGGCACTCCTCAGTAGTGGT 762
Qy 263 AsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyValThrProLys----- 280
Db 763 ATTGACGCCCTCTGCTCCAA-----GCACACCCGAGCTGGACT 801
Qy 281 ProSerLeuLeuLysAlaAlaLeuIleAlaGlyAla----- 292
Db 802 CCAGACAAAGTAAACAGCCCTCATAGAACTGCTGATATCGTAAAGCCAGATGAATA 861
Qy 293 AlaAspValGlyLeuGlyPheProAsnGlyAsnGlnGlyTyrGlyArgValThrLeuAsp 312
Db 862 GCCGATATAGCTACGGTGCA-----GGTAGGGTTAATGCATAC 900
Qy 313 LysSerLeu-----AsnValAlaPheValAsnGluThrSerProLeuSerThrSer 329
Db 901 AAGGTATAAATACTACGATACTATGCAAGAGCTAGTGTTCACCTGATATGTTGCCAACA 960
Qy 330 GlnLysAlaThrTyrSerPheThrAlaGlnAlaGlyLysProLeuLysIleSerLeuVal 349
Db 961 GGCAGCCAACTCACAGTTCGTATATAGCGAGAGCTGCTGCTGATGCTGACCATTTATAC 1020
Qy 350 TrpSerAspAlaProGlySerThrThrAlaSerLeuThrLeuValAsnAspLeuAspLeu 369
Db 1021 TGGGCAATGCAAT-----AGCGACCTTGATCTT 1050
Qy 370 VallIleThrAlaProAsnGlyThrLysTyrValGlyAsnAspPheThrAlaProTyrAsp 389
Db 1051 TACCTCTACGATCCCAATGGAACACAG---GTTGACTACTCTTACACCGCTACTAT--- 1104
Qy 390 AsnAsnTrpAspGlyArgAsnAsnValGluAsnValPheIleAsnAlaProGlnSerGly 409
Db 1105 -----GGATTGCAAAAGTGTGTTATTACAAACCACTGATGGA 1143
Qy 410 ThrTyrThrValGluValGlnAlaTyrAsn 419
Db 1144 ACATGGACAATTAAGGTTGTAAGCTACAGC 1173
RESULT 4
US-10-090-624-15
```

```
; Sequence 15, Application US/10090624
; Publication No. US20020132335A1
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: SHIMOJO, Tomoko
; APPLICANT: ASADA, Kiyozo
; APPLICANT: KATO, Ikunoshi
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
; FILE REFERENCE: TAKAKURA-6
; CURRENT APPLICATION NUMBER: US/10/090,624
; CURRENT FILING DATE: 2002-03-06
; PRIOR APPLICATION NUMBER: 09/445,472
; PRIOR FILING DATE: 1999-12-06
; PRIOR APPLICATION NUMBER: 151969/1997
; PRIOR FILING DATE: 1997-06-10
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 15
; LENGTH: 1962
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
US-10-090-624-15
Alignment Scores:
Pred. No.: 6,27e-33 Length: 1962
Score: 396.00 Matches: 127
Percent Similarity: 42.56% Conservative: 56
Best Local Similarity: 29.53% Mismatches: 163
Query Match: 17.58% Indels: 84
DB: 15 Gaps: 15
US-09-985-689a-7 (1-433) x US-10-090-624-15 (1-1962)
Qy 16 AsnPheGlyLeuTyrGlyGlnGlyIleValAlaValAlaAspThrGlyLeuAspThr 35
Db 454 AACTGGGATATGATGGTTCCTGGATCAATAGGAATAATTGACATGGAAATTCAC-- 510
Qy 36 GlyArgAsnAspSerSerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAla 55
Db 511 -----GCTCTCATCCAGATCTCCAAAGGAAAGTA----- 540
Qy 56 LeuGlyArgThrAsnAsnAlaAsn-----AspProAsnGlyHisGly 69
Db 541 ATGGTGGGTAGATTTGTCTAGTGGAGTATCCATACATGACCATGGACATGGA 600
Qy 70 ThrHisValAlaGlySerValLeuGly-----AsnAlaThrAsnLysGly 84
Db 601 ACTCATGTAGCTTCAATAGCAGCTGCTACTGGAGCAGCAAGTAATGGCAAGTACAGGA 660
Qy 85 MetAlaProGlnAlaAsnLeuValPheGlnSerIleMet-----AspSerGlyGly 102
Db 661 ATGGCTCCAGGAGCTAAGCTGGCGGAATTAAGTTCCTAGTGGCGATGTTCTGGAAGC 720
Qy 103 LeuGlyGlyLeuProAlaAsnLeuGlnThrLeuPheSerGlnAlaTyrSerAlaGlyAla 122
Db 721 ATATCTACTATAATTAAGGAGTGTGAGTGGCGCGTGTGATAACAAGATAAGTACGGAAT 780
Qy 123 ArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyrThrThrAspSerArg 142
Db 781 AAGGTCAATTAATCTTTCTTTGGTTCAAGCCAGCTCAGATGTTGACGCTCTAAGT 840
Qy 143 AsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAlaGlyAsnGlu 162
Db 841 CAGGCTGTTAATGTCAGCGTGGATGCTGATTAAGTGTGCTGGCGTGGAAACAGT 900
Qy 163 GlyProGlySerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrValGly 182
Db 901 GGACCTAACAGTATACAACTGGTCTCCAGCGAGCTGCAAGCAAGTTATTACAGTTGA 960
Qy 183 AlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsnHisValAla 202
```

Db 961 GCC-----GTTGCAAGTATGATGTTATAACA 987  
Qy 203 GlnPheSerSerArgGlyProThrArgAspGlyArgIleLeuProAspValMetAlaPro 222  
Db 988 AGCTTCTCAAGCAGAGGGCCAACTGCAGACGGCCTTAAGCTTAAAGCTTCTGCTCCA 1047  
Qy 223 GlyThrTyrIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrpAlaAsn 242  
Db 1048 GGAACCTGATTAATCTCCAGAGCAGT-----GGAACCTAGCATGGTCAACCA 1098  
Qy 243 HisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIleValAlaGly 262  
Db 1099 ATTAATGACTATTACACAGCAGCTCTCGGACATCAATGGCACTCTCCACGTACTGGT 1158  
Qy 263 AsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyValThrProLys----- 280  
Db 1159 ATTGCAGCCCTCTGCTCCAA-----GCACACCCGAGCTGGACT 1197  
Qy 281 ProSerLeuLysAlaAlaLeuLeuAlaGlyAla----- 292  
Db 1198 CCAGACAAAGTAAACACACCTCATAGAACTGCTGATATCGTAAAGCCAGATGAATA 1257  
Qy 293 AlaAspValGlyLeuGlyPheProAsnGlyAsnGlnGlyTyrGlyArgValThrLeuAsp 312  
Db 1258 GCCGATATAGCCTACGCTGCA-----GGTAGGTTAATGCATAC 1296  
Qy 313 LysSerLeu-----AsnValAlaPheValAsnGluThrSerProLeuSerThrSer 329  
Db 1297 AGGCTATATAACATACATATGCAAGCTAGTGTTCACCTGGATATGTTGCCAACAA 1356  
Qy 330 GlnLysAlaThrTyrSerPheThrAlaGlnAlaGlyLysProLeuLysIleSerLeuVal 349  
Db 1357 GGCAGCAAACTCACCAGTTCGTATTAGCGAGCTTCGTTAACTGCCACATTATAC 1416  
Qy 350 TrpSerAspAlaProGlySerThrThrAlaSerLeuThrLeuValAsnAspLeuAspLeu 369  
Db 1417 TGGACAAATGCCAAT-----AGCACCTTCATCTT 1446  
Qy 370 ValIleThrAlaProAsnGlyThrLysTyrValGlyAsnAspPheThrAlaProTyrAsp 389  
Db 1447 TACCTCTACGATCCCAATGAAACAG--GTTGACTACTCTTACACCGCTACTAT--- 1500  
Qy 390 AsnAsnTrpAspGlyArgAsnAsnValGluAsnValPheIleAsnAlaProGlnSerGly 409  
Db 1501 -----GGATTGAAAGGTTGTTATTACAAACCACTGATGGA 1539  
Qy 410 ThrTyrThrValGluValGlnAlaTyrAsn 419  
Db 1540 ACATGGACAAATTAAAGTTCTAAGCTACAGC 1569

RESULT 5  
US-10-156-761-5701  
Sequence 5701, Application US/10156761  
Publication No. US20030119018A1  
GENERAL INFORMATION:  
APPLICANT: OMURA, SATOSHI  
APPLICANT: IKEDA, HARUO  
APPLICANT: ISHIKAWA, JUN  
APPLICANT: HORIKAWA, HIROSHI  
APPLICANT: SHIBA, TADAYOSHI  
APPLICANT: SAKAKI, YOSHIYUKI  
APPLICANT: HATTORI, MASAHIRA  
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
FILE REFERENCE: 249-262  
CURRENT APPLICATION NUMBER: US/10/156,761  
CURRENT FILING DATE: 2002-05-29  
PRIOR APPLICATION NUMBER: JP 2001-204089  
PRIOR FILING DATE: 2001-05-30  
PRIOR APPLICATION NUMBER: JP 2001-272697  
PRIOR FILING DATE: 2001-08-02  
NUMBER OF SEQ ID NOS: 15109  
SEQ ID NO 5701

LENGTH: 3624  
TYPE: DNA  
ORGANISM: Streptomyces avermitilis  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (1)..(3624)  
US-10-156-761-5701  
Alignment Scores:  
Pred. NO.: 4,49e-28 Length: 3624  
Score: 355.00 Matches: 126  
Percent Similarity: 44.47% Conservative: 51  
Best Local Similarity: 31.66% Mismatches: 151  
Query Match: 15.76% Indels: 70  
DB: 14 Gaps: 17  
US-09-985-689A-7 (1-433) x US-10-156-761-5701 (1-3624)  
Qy 8 ValLysAlaAspValAlaGlnAsnAsn-----Phe 17  
Db 556 GTGAGGCCGACATGCGCGAGAGACGACGCGACATCGTACGCGCGCGGTGGAGGCC 615  
Qy 18 GlyLeuTyrGlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArg 37  
Db 616 GGGTCTCACGGCGCAGCGCGTCACCGCTCGCGCTGTCGACACCGCGCGTGCACAC-- 669  
Qy 38 AsnAspSerSerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGly 57  
Db 670 -----ACTCACCCGACCTCGCGCGCGGTGTCCCGGAGCAAGAGCTTCATC 717  
Qy 58 ArgThrAsnAsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeu 77  
Db 718 GACGGGAGGAGGTGTCGCGACCGCAACGCGCACGGGACCCACGTCACCTCGACCGTCGC 777  
Qy 78 Gly-----AsnAlaThrAsnLysGlyMetAlaProGlnAlaAsnLeuVal 92  
Db 778 GGCAGCGCGCGCGCTCCGACGCGACGCGCGCGGTGCGCGCGGTGCGCGCTCGCC 837  
Qy 93 PheGlnSerIleMet---AspSerGlyGlyGlyLeuGlyGlyLeuProAlaAsnLeuGln 111  
Db 838 GTCGCAAGTGTCTACGCGACCGACCGAGGCGCGGGA-----ACGAGTCCCAG 882  
Qy 112 ThrLeuPheSerGlnAlaTyrSerAla-----GlyAlaArgIleHisThrAsnSer 128  
Db 883 ATCATCGCGCGCATGGAATGCGCGCGCGGACGTCGTCGCGAGGATCGTCTCGATGAGC 942  
Qy 129 TrpGlyAlaProValAsnGlyAlaTyrThrThrAsp-----SerArgAsnValAspAsp 146  
Db 943 CTCGGATCG--ACCGAGGCGCAGCGAGCGGACCGACCCCATGCCCGCGCGCTCGACACC 999  
Qy 147 TyrValArgLysAsnAspMetThrIleLeuPheAlaAlaGlyAsnGluGlyProGlySer 166  
Db 1000 CTCTCCGAGGAGACCGCGCGCTCTTCGTCGCGCGGGAACACCGGTGCGCCCTCC 1059  
Qy 167 GlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsn 186  
Db 1060 ---TCGATCGGCTCGCGCGCGCGCGCTCGCGCTGACCTGCGCGCGCGCTCC 1107  
Qy 187 LeuArgProSerPheGlySerTyrAlaAspAsnIleAsnHisValAlaGlnPheSerSer 206  
Db 1108 -----GTGACTCATCCGACCGCGCGCTACTTCCACGAGC 1143  
Qy 207 ArgGlyProThrArg---AspGlyArgIleLysProAspValMetAlaProGlyThrTyr 225  
Db 1144 GCGCGCGCGCGCGCACGCGCGACAGCCCTCAAGCCGACCTCGCCGACCGCGCTCGAC 1203  
Qy 226 IleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrpAlaAsnHisAspSer 245  
Db 1204 ATCCGCGCGCGCGCTCCCAGCTCGCGCGCGCGCGCGCGCTAC----- 1245  
Qy 246 LysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIleValAlaGlyAsnValAla 265  
Db 1246 ---TACACCTCCATGAGCGGTACGTCGATGCGACGCCCATCGCGGGGTGCGCGCG 1302

QY 266 GlnLeuArgGluHisPheValysAsnArgGlyValThrProLysProSerLeuLeuLys 285  
 Db 1303 CTCCTCCGCGAGCAGACCCCGACTGAGCCGGCGCGGCTCAAGACCGCTGATGTC 1362  
 QY 286 AlaAla-----LeuIleAlaGlyAlaAlaAspValGlyLeuGlyPheProAsnGlyAsn 303  
 Db 1363 ACGTCCGAGCAACTCGACGCTCCGTATATACAGTTGGGGCGGGTCCGGTCAAGTGTGCGG 1422  
 QY 304 GlnGlyTrpGly---ArgValThrLeuAspLysSerLeuAsnValAlaPhe----- 319  
 Db 1423 GACGCCGTCCGCGCGCCCGTCAACCGCAGCGCGCGCGCGCTCCACCCCTCCACCCCTGG 1482  
 QY 320 -----ValAsnGluThrSerProLeuSerThrSerGlnLysAla 332  
 Db 1483 CCCCATGACGCGGATGACCCGCTCAGACAGCGGTCACTTCTCAACTCTCCGACAG 1542  
 QY 333 ThrTyrSerPheThr---AlaGlnAlaGlyLysProLeuLysLeuSerLeuValTrpSer 351  
 Db 1543 ACGTCCGAGTTGAGCCTCGCGGTGCGGGCGCGCGCGCGCGCGGTGCGCCACCTCGCGGAC 1602  
 QY 352 AspAla-----ProGlySerThrThrAlaSerLeuThrLeuValAsnAsp 366  
 Db 1603 ACCGCATCCACGTCGCGCGCCCGCCAGCGCGCGCGCGCCACCCGTCGACCGCGCGAC 1656

## RESULT 6

US-10-156-761-1  
 ; Sequence 1, Application US/10156761  
 ; Publication No. US20030119018A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: OMURA, SATOSHI  
 ; APPLICANT: IKEDA, HARUO  
 ; APPLICANT: ISHIKAWA, JUN  
 ; APPLICANT: HORIKAWA, HIROSHI  
 ; APPLICANT: SHIBA, TADAYOSHI  
 ; APPLICANT: SAKAKI, YOSHIYUKI  
 ; APPLICANT: HATTORI, MASAHIRA  
 ; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
 ; FILE REFERENCE: 249-262  
 ; CURRENT APPLICATION NUMBER: US/10/156,761  
 ; CURRENT FILING DATE: 2002-05-29  
 ; PRIOR APPLICATION NUMBER: JP 2001-204089  
 ; PRIOR FILING DATE: 2001-05-30  
 ; PRIOR APPLICATION NUMBER: JP 2001-272697  
 ; PRIOR FILING DATE: 2001-08-02  
 ; NUMBER OF SEQ ID NOS: 15109  
 ; SEQ ID NO 1  
 ; LENGTH: 9025608  
 ; TYPE: DNA  
 ; ORGANISM: Streptomyces avermitilis  
 ; FEATURE:  
 ; NAME/KEY: misc feature  
 ; LOCATION: (4187715)  
 ; OTHER INFORMATION: a, t, c, g, other or unknown  
 US-10-156-761-1

Alignment Scores:  
 Pred. No.: 3,47e-23 Length: 9025608  
 Score: 355.00 Matches: 126  
 Percent Similarity: 44.47% Conservative: 51  
 Best Local Similarity: 31.66% Mismatches: 151  
 Query Match: 15.76% Indels: 70  
 DB: 14 Gaps: 17

US-09-985-689A-7 (1-433) x US-10-156-761-1 (1-9025608)

QY 8 ValLysAlaAspValAlaGlnAsnAsn-----Phe 17  
 Db 6918813 GTCGAGCGCATGTCGCGGAGACGACCGCGAGATCGGTACCGCGCGCGCTGGGACGCC 6918872  
 QY 18 GlyLeuTyrGlyGlnGlyGlnIleValAlaAlaAspThrGlyLeuAspThrGlyArg 37  
 Db 6918873 GGCGTCACGGCGCAGCGCGCTCACCGTCGCGCGTGTCTCGACACCGCGCGGTTCGACACC----- 6918926

QY 38 AsnAspSerSerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGly 57  
 Db 6918927 -----ACTCACCCCGACCTCCGCGCGCGGTGTCCCGAGCAAGACTTCATC 6918974  
 QY 58 ArgThrAsnAsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeu 77  
 Db 6918975 GACGGGAGGAGGTGCGCGACCGCAACCGCCACCGGACCCACGTCACCTCGACCGTGGC 6919034  
 QY 78 Gly-----AsnAlaThrAsnLysGlyMetAlaProGlnAlaAsnLeuVal 92  
 Db 6919035 GGCAGCGCGCGCTCCGACGCGCAGCGCGCGCTCGCGCGCGGTGCGCGCTCGCC 6919094  
 QY 93 PheGlnSerIleMet---AspSerGlyGlyLeuGlyGlyLeuProAlaAsnLeuGln 111  
 Db 6919095 CTCGCAAGTGTCTCAGCAGCAGCGCGCGGGA-----AGCGAGTCCGAG 6919139  
 QY 112 ThrLeuPheSerGlnAlaTyrSerAla-----GlyAlaArgIleHisThrAsnSer 128  
 Db 6919140 ATCATCGCGGCGCATGGAATGGCGCGCGGACGTGGTGTGCGAGATCGTCTCGATGAGC 6919199  
 QY 129 TrpGlyAlaProValAsnGlyAlaTyrThrThrAsp-----SerArgAsnValAspAsp 146  
 Db 6919200 CTCGGATCG---ACCGAGCGCAGCGACCGGACCCCATGGCGGCGCGTCCGACACC 6919256  
 QY 147 TyrValArgLysAsnAspMetThrIleLeuPheAlaIleAlaGlyAsnGluGlyProGlySer 166  
 Db 6919257 CTCCTCGAGGACCGCGCGCTCTCTGCTGCGCGGGGAAACACCGGTGCCCCCTCC 6919316  
 QY 167 GlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsn 186  
 Db 6919317 ---TCGATCGGCTCGCGCGCGCGCGCGCTCCGCTGACCGTGGCGGCGC----- 6919364  
 QY 187 LeuArgProSerPheGlySerTyrAlaAspAsnIleAsnHisValAlaGlnPheSerSer 206  
 Db 6919365 -----GTCGACTCATCTCCGACCGCGCGCTTACTTCCACGAG 6919400  
 QY 207 ArgGlyProThrArg---AspGlyArgIleLysProAspValMetAlaProGlyThrTyr 225  
 Db 6919401 GCGCGCGCGCGCACGCGCACAAACCGCTCAAGCGCGCGCTCCGCGCGCGCGGTGCGAC 6919460  
 QY 226 IleLeuSerAlaArgSerSerLeuAlaProAspSerPheTrpAlaAsnHisAspSer 245  
 Db 6919461 ATCCGCGCGCGCGCTCCGAGCTCCGCGCGCGCGCGCGCGCGCGCTAG----- 6919502  
 QY 246 LysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIleValAlaGlyAsnValAla 265  
 Db 6919503 ---TACACCTCCATGAGCGGTACGTGATGGCGACGCGCGCTCGCGGGTCCGCGCG 6919559  
 QY 266 GlnLeuArgGluHisPheValLysAsnArgGlyValThrProLysProSerLeuLys 285  
 Db 6919560 CTCCTCGCGGAGCAGCAGCCCGACTGGACCGCGCGCGCGCTCAAGACCGCTGATGTC 6919619  
 QY 286 AlaAla-----LeuIleAlaGlyAlaAlaAspValGlyLeuGlyPheProAsnGlyAsn 303  
 Db 6919620 ACGTCCGAGCACTCCGACGCTCCGATATATCAGTTGGGGCGGTGCGGTGAGTGTGCGG 6919679  
 QY 304 GlnGlyTrpGly---ArgValThrLeuAspLysSerLeuAsnValAlaPhe----- 319  
 Db 6919680 GACCGCGTCCGCGCGCGCGCTCACCGCGCGCGCGCGCGCTCCGCTTCCACCGCTGG 6919739  
 QY 320 -----ValAsnGluThrSerProLeuSerThrSerGlnLysAla 332  
 Db 6919740 CCCCATGACCGCGATGACCGCGTCCAGACGCGTCACTTCTCAACTCTCCGACAG 6919799  
 QY 333 ThrTyrSerPheThr---AlaGlnAlaGlyLysProLeuLysLeuSerLeuValTrpSer 351  
 Db 6919800 ACGTCCGAGTTGAGCCTCGCGGTGCGCGCGCGCGCGCGGTGTCGCCACCTTCGCGGAC 6919859  
 QY 352 AspAla-----ProGlySerThrThrAlaSerLeuThrLeuValAsnAsp 366  
 Db 6919860 ACGGCATCCAGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGGTGACCGCGGAC 6919913



```
RESULT 7
US-10-156-761-3306
; Sequence 3306, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 3306
; LENGTH: 3417
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURES:
; NAME/KEY: CDS
; LOCATION: (1)..(3417)
US-10-156-761-3306

Alignment Scores:
Pred. No.: 1,33e-25 Length: 3417
Score: 332.00 Matches: 137
Percent Similarity: 44.39% Conservative: 53
Best Local Similarity: 32.01% Mismatches: 161
Query Match: 14.74% Indels: 78
DB: 20 Gaps: 20

US-09-985-689a-7 (1-433) x US-10-156-761-3306 (1-3417)
Qy 18 GlyLeuTyrglyGinglyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyAyg 37
Db 745 GGGTACGACGCGCAAGGGGCTCAAGTCCGCTCTGACACCGGTGTCGAC----- 795
Qy 38 AsnAspSerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyAlaLeuGly 57
Db 796 -----GGACCCACCGCGACCTCAAGGACCGAGTGTGCGCGGCTCAAGAACTTCTCC 846
Qy 58 ArgThrAsnAlaAsnAspProAsnGlyHisIleThrHisValAlaGlySerValLeu 77
Db 847 GCCGCCCGCGACGCGCGGACCACTTCGGTACCGGACGCGACGCGTCCATCGCGGCG 906
Qy 78 Gly-----AsnAlaThrAsnIlysglyMetAlaProGlnAlaAsnLeuVal 92
Db 907 GGCACCGCGCGCAAGTCCCAAGCGCAAGTACAGGGTGTGCGCGCGCGCGGACGATCCTC 966
Qy 93 PheGlnSerIleMetAsp-----SerGlyGlyLeuGlyLeuGlyLeuProAlaAsnLeu 110
Db 967 AACGGCAAGGTCTCGACGACACCGGCTCGCGGACGACTCCGCGCATCTCGCGCGCATG 1026
Qy 111 GlnThrLeuPheSerGlnAlaTySerAlaGlyAlaArgIleHisThrAsnSerTrpGly 130
Db 1027 GAGTGGGGCGCGGACGAG-----GGCGCCGCGCTCGTCACTCGAGCTCGGCG 1074
Qy 131 AlaProValAsnGlyAlaTyThrThrAspSerArgAsnValAspAspTyValArgLys 150
Db 1075 -----GGCGGCGACACCGCGGACGCTCGGAGCGGAGGTCAACAG 1122
Qy 151 -----AsnAspMetThrIleLeuPhe-----AlaAlaGlyAsnGlyPro---Gly 165
Db 1123 CTCTCCGAGGAGAGGGGATCTCTTCCGATCGCGCGCGGCAACGAGGCGGATTCGGC 1182
Qy 156 SerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrValGlyAlaThrGlu 185
```

## RESULT 8

```
US-10-156-761-1/c
; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
```

## RESULT 8

```
US-10-156-761-1/c
; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
```

; PRIOR APPLICATION NUMBER: JP 2001-272697  
 ; PRIOR FILING DATE: 2001-08-02  
 ; NUMBER OF SEQ ID NOS: 15109  
 ; SEQ ID NO 1  
 ; LENGTH: 9025608  
 ; TYPE: DNA  
 ; ORGANISM: Streptomyces avermitilis  
 ; FEATURE:  
 ; NAME/KEY: misc feature  
 ; LOCATION: (4187715)  
 ; OTHER INFORMATION: a, t, c, g, other or unknown  
 US-10-156-761-1

Alignment Scores:  
 Pred. No.: 1,12e-20 Length: 9025608  
 Score: 332.00 Matches: 137  
 Percent Similarity: 44.39% Conservative: 53  
 Best Local Similarity: 32.01% Mismatches: 161  
 Query Match: 14.74% Indels: 78  
 DB: 14 Gaps: 20

US-09-985-689A-7 (1-433) x US-10-156-761-1 (1-9025608)

Qy 18 GlyLeuTyGlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArg 37  
 Db 4132505 GGATACGACGGAGGCGGATCAAGATCGCGTCCGTCGACACCGGTCGAC----- 4132455  
 Qy 38 AsnAspSerSerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyAlaLeuGly 57  
 Db 4132454 -----GCGACCCACCGGACCTCAAGACCGAGTGGCGGAGTCCAGAACTTCTCC 4132404  
 Qy 58 ArgThrAsnAsnAlaAspProAsnGlyHisGlyThrHisValAlaGlySerValLeu 77  
 Db 4132403 GCGCGCGCGACGCGCCGACCACTTCGTCGACGACGACGTCGCTCCATCGCGCG 4132344  
 Qy 78 Gly-----AsnAlaThrAsnLysGlyMetAlaProGlnAlaAsnLeuVal 92  
 Db 4132343 GGCACCGCGCGCCCAAGTCAACGCAAGTACAAAGGTCGTGCGCGCGCGGCGACGATCTC 4132284  
 Qy 93 PheGlnSerIleMetAsp-----SerGlyGlyGlyLeuGlyLeuProAlaAsnLeu 110  
 Db 4132283 AACGGCAAGGTCTTCGACACACCGCGCTCGCGGACGACTCGGCATCTCGCGGCGATG 4132224  
 Qy 111 GlnThrLeuPheSerGlnAlaTySerAlaGlyAlaArgIleHisThrAsnSerTrpGly 130  
 Db 4132223 GAGTGGCGCGCGCGAGCAG-----GGCGCGGACGTCGTCACACTGAGCTCGGC 4132176  
 Qy 131 AlaProValAsnGlyValaTyThrThrAspSerArgAsnValaAspTyValArgLys 150  
 Db 4132175 -----GGCGCGACACCCCGGAGATCGACCGCTGGAGCGGAGTCAACAG 4132128  
 Qy 151 -----AsnAspMetThrIleLeuPhe-----AlaAlaGlyAsnGlyPro---Gly 165  
 Db 4132127 CTCTCGGAGGAGGAGGCGATCTCTTCGCGATCGCGCGCGGCAACGAGGCGAGTTCGGC 4132068  
 Qy 166 SerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrValGlyValaThrGlu 185  
 Db 4132067 GAGCAGACCATCGCTCCCGGCGGCGGCGGCGGCGCTACCGCTCGGCGCC----- 4132014  
 Qy 186 AsnLeuArgProSerPheGlySerTyAlaAspAsnIleAsnHisValAlaGlnPheSer 205  
 Db 4132013 -----GTGAACGACGACGACGACGAGTCCGCTCTCTCC 4131981  
 Qy 206 SerArgGlyProThrArgAspGlyArgIleLysProAspValMetAlaProGlyThrTy 225  
 Db 4131980 AGCGCGGCGCGCGCTGACGCGCGCATCAAGCCGCGAGTCAACGACCGCGGCGTGCAC 4131921  
 Qy 226 IleLeuSerAlaArgSerSerLeuAlaProAspSerPheTrpAlaAsnHisAspSer 245  
 Db 4131920 ATCACCAGCGCGC-----GGCGCGCGCGGCGGCGGCGTCAACGAGGAGTGGCGAG 4131870  
 Qy 246 Lys-----TyrAlaTyMetGlyGlyThrSerMetAlaThrProIleValAlaGly 262

Db 4131869 AAGCGGACGGCTACCTACCATCTCGGTACGTGCGACCGCGCATGTGCGGGC 4131810  
 Qy 263 AsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyValThrProLysProSer 282  
 Db 4131809 GCGCGCGGATCTCTCAAGCAGCAGCAG-----CCCAACTGGTCG 4131771  
 Qy 283 Leu-----LeuLysAlaAlaLeuIleAlaGlyAlaAlaAspValGlyLeuGlyPhePro 300  
 Db 4131770 TTCGCGGAGCTCAAGGGCGCGCTG---ACGGCTCCCGAGGGCGGCAAGTACACGCCG 4131714  
 Qy 301 AsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaPheVal 320  
 Db 4131713 ---TTCCAGCAGGCGCTCGGCGGTATCGCGTGCACAAAGCGCATCAAGCATCTCGGTATC 4131657  
 Qy 321 AsnGluThrSerProLeuSerThrSerGlnLysAlaThrTyThrPheThrAlaGlnAl 340  
 Db 4131656 -----GCCACCCGGAAGTCTGGTG-----AGCTCGGCATCCAGCAG 4131621  
 Qy 340 aGlyLysProLeuLysIleSerLeuValTrpSerAspAlaProGlySerThrThrAla 360  
 Db 4131620 TGGCGCGCACCGCAGCAGCAAGCGGTCAACGAGCAGCTACCTACCGCAACCTCGGCGAG 4131561  
 Qy 360 rLeuThrLeuValAsnAspLeuValIleThrAlaProAsnGlyThr----- 377  
 Db 4131560 AGTGACGTC-ACGTGAACCTCGGTGCGCGGCGACCAACCCCAAGGCGTCCGCGCTCC 4131502  
 Qy 378 -----LysTyValGlyAsnAspPheThrAlaProTyThrAspAsnAsnTrpAs 393  
 Db 4131501 GTCGCGCTTCTTCAAGCTCGCGCGCAGCAAGGTCAACGCTCCCGCGC-----GG 4131454  
 Qy 393 pGlyArgAsnAsnValGlu-----AsnValPheIleAsnAlaProGlnSerGlyTh 410  
 Db 4131453 CGCGAAGGCTCGGTGCGTCTCAGGTCAACGAGTGGCGGCGCACCGACGCGCGC 4131394  
 Qy 410 rTyThrValGluValGlnAla 417  
 Db 4131393 GTACTCGCGTACGTGACGCGC 4131372

RESULT 9  
 US-09-974-300-1934  
 ; Sequence 1934 Application US/09974300  
 ; Patent No. US20020146721A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Berka, Randy M.  
 ; APPLICANT: Clausen, ID Groth  
 ; TITLE OF INVENTION: Methods For Monitoring Multiple Gene  
 ; TITLE OF INVENTION: Expression  
 ; FILE REFERENCE: 10085.500-US  
 ; CURRENT APPLICATION NUMBER: US/09/974,300  
 ; CURRENT FILING DATE: 2001-10-05  
 ; PRIOR FILING DATE: 2000-10-06  
 ; PRIOR APPLICATION NUMBER: 09/580,598  
 ; PRIOR FILING DATE: 2001-03-27  
 ; NUMBER OF SEQ ID NOS: 8481  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 1934  
 ; LENGTH: 1329  
 ; TYPE: DNA  
 ; ORGANISM: Bacillus licheniformis  
 US-09-974-300-1934

Alignment Scores:  
 Pred. No.: 1.98e-25 Length: 1329  
 Score: 325.00 Matches: 105  
 Percent Similarity: 43.98% Conservative: 41  
 Best Local Similarity: 31.63% Mismatches: 118  
 Query Match: 14.43% Indels: 68  
 DB: 9 Gaps: 13

US-09-985-689A-7 (1-433) x US-09-974-300-1934 (1-1329)

Qy 11 AspValAlaGlnAsnAsnPheGlyLeuTyGlyGlnIleValAlaValAlaAsp 30

```

; APPLICANT: CHENG, Yi-Qiang
; APPLICANT: TANG, Gong-Li
; TITLE OF INVENTION: Discrete Acyltransferases Associated with Type I Polyketide
; FILE REFERENCE: 054030-0021
; CURRENT APPLICATION NUMBER: US/10/314,657
; CURRENT FILING DATE: 2002-12-09
; PRIOR APPLICATION NUMBER: PCT/US02/08937
; PRIOR FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: US 60/278,935
; PRIOR FILING DATE: 2001-03-26
; NUMBER OF SEQ ID NOS: 214
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
; LENGTH: 135638
; TYPE: DNA
; ORGANISM: Streptomyces atroolivaceus
; US-10-314-657-1

Alignment Scores:
Pred. No.: 2,55e-22 Length: 135638
Score: 323.00 Matches: 133
Percent Similarity: 41.63% Conservative: 56
Best Local Similarity: 29.30% Mismatches: 137
Query Match: 14.34% Indels: 128
DB: 14 Gaps: 22

US-09-985-689A-7 (1-433) x US-10-314-657-1 (1-135638)
QY 6 GlyLeuValValAlaAspValAlaGlnAsnAsnPhe----- 17
DB 8067 GCGAGGTTGAAGCGCATCTGCGCGACTCCACCCGCCAGATCGCGCGCGAGAAAGGTATGG 8126
QY 18 -----GlyLeuTyrGlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThr 35
DB 8127 CGGAGGCGGACACCGGCCAGAGGTGAGTTCGCGATGTCGACAGCGGCGCGACAC 8186
QY 36 GlyArgAsnAspSerSerMetHisGluAlaPheArgGlyLysIle-----ThrAlaLeu 53
DB 8187 -----GAACACCGCGACCTGTCGCGCAGAGTGTCCGACAGCGCGCAGC 8228
QY 54 TyrAlaLeuGlyArgThrAsnAsnAlaAsnAspProAsnGlyHisGlyThrHisValAla 73
DB 8229 TTGCTGCCCGCGGAGGACGATCGCC---GACTACACGCGCCAGCGACGACGTCGCC 8295
QY 74 GlySerValLeuGlyAsnAlaThr-----AsnLysGlyMetAlaProGln 88
DB 8286 TCGACCATGTCGCGCAGCGGCGCCTCCGACGCGAGGAGCGGGTGTGCGCTCCGCGC 8345
QY 89 AlaAsnLeuValPheGlnSerIleMetAspSerGlyGlyGly----- 102
DB 8346 GCGCGGCTGTCGTCGCGCAAGGTGCTCACTCCGAGGCGCGCGCCAGGATCTGGATC 8405
QY 103 LeuGlyGlyLeuProAlaAsnLeuGlnThrLeuPheSerGlnAlaTyrSerAlaGlyAla 122
DB 8406 ATCGCGGCGCATGGAG-----TGGCGCGCGCGCGACGACGAGGCGC 8444
QY 123 ArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyrThrAsp----- 140
DB 8445 AGGATCATCAGCATGAGCTGGGC-----GGCGCGCGGTGACAGAGACCGCCGATG 8495
QY 141 SerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaGly 160
DB 8496 AGCGAGCGCTGCGAAGTCACTAGCCAGCACGCGCGCGGTGTTGTCGTGATCGCGCGGC 8555
QY 161 AsnGluGlyProGlySerGlyThrIleSerAlaProGlyThrAlaLysAsnLallethr 180
DB 8556 AACGCGCGCGCGCACTCC-----ATCAGCAGCGCGGTGCGGCGAGACTCGCGCTGACC 8609
QY 181 ValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsnHis 200
DB 8610 GTGCGCGCC-----GTGCACTCCACCGACAGC 8636

; RESULT 10
; US-10-314-657-1
; Sequence 1, Application US/10314657
; Publication No. US2003017588A1
; GENERAL INFORMATION:
; APPLICANT: SHEN, Ben
```

```

; APPLICANT: CHENG, Yi-Qiang
; APPLICANT: TANG, Gong-Li
; TITLE OF INVENTION: Discrete Acyltransferases Associated with Type I Polyketide
; FILE REFERENCE: 054030-0021
; CURRENT APPLICATION NUMBER: US/10/314,657
; CURRENT FILING DATE: 2002-12-09
; PRIOR APPLICATION NUMBER: PCT/US02/08937
; PRIOR FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: US 60/278,935
; PRIOR FILING DATE: 2001-03-26
; NUMBER OF SEQ ID NOS: 214
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
; LENGTH: 135638
; TYPE: DNA
; ORGANISM: Streptomyces atroolivaceus
; US-10-314-657-1

Alignment Scores:
Pred. No.: 2,55e-22 Length: 135638
Score: 323.00 Matches: 133
Percent Similarity: 41.63% Conservative: 56
Best Local Similarity: 29.30% Mismatches: 137
Query Match: 14.34% Indels: 128
DB: 14 Gaps: 22

US-09-985-689A-7 (1-433) x US-10-314-657-1 (1-135638)
QY 6 GlyLeuValValAlaAspValAlaGlnAsnAsnPhe----- 17
DB 8067 GCGAGGTTGAAGCGCATCTGCGCGACTCCACCCGCCAGATCGCGCGCGAGAAAGGTATGG 8126
QY 18 -----GlyLeuTyrGlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThr 35
DB 8127 CGGAGGCGGACACCGGCCAGAGGTGAGTTCGCGATGTCGACAGCGGCGCGACAC 8186
QY 36 GlyArgAsnAspSerSerMetHisGluAlaPheArgGlyLysIle-----ThrAlaLeu 53
DB 8187 -----GAACACCGCGACCTGTCGCGCAGAGTGTCCGACAGCGCGCAGC 8228
QY 54 TyrAlaLeuGlyArgThrAsnAsnAlaAsnAspProAsnGlyHisGlyThrHisValAla 73
DB 8229 TTGCTGCCCGCGGAGGACGATCGCC---GACTACACGCGCCAGCGACGACGTCGCC 8295
QY 74 GlySerValLeuGlyAsnAlaThr-----AsnLysGlyMetAlaProGln 88
DB 8286 TCGACCATGTCGCGCAGCGGCGCCTCCGACGCGAGGAGCGGGTGTGCGCTCCGCGC 8345
QY 89 AlaAsnLeuValPheGlnSerIleMetAspSerGlyGlyGly----- 102
DB 8346 GCGCGGCTGTCGTCGCGCAAGGTGCTCACTCCGAGGCGCGCGCCAGGATCTGGATC 8405
QY 103 LeuGlyGlyLeuProAlaAsnLeuGlnThrLeuPheSerGlnAlaTyrSerAlaGlyAla 122
DB 8406 ATCGCGGCGCATGGAG-----TGGCGCGCGCGCGACGACGAGGCGC 8444
QY 123 ArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyrThrAsp----- 140
DB 8445 AGGATCATCAGCATGAGCTGGGC-----GGCGCGCGGTGACAGAGACCGCCGATG 8495
QY 141 SerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaGly 160
DB 8496 AGCGAGCGCTGCGAAGTCACTAGCCAGCACGCGCGCGGTGTTGTCGTGATCGCGCGGC 8555
QY 161 AsnGluGlyProGlySerGlyThrIleSerAlaProGlyThrAlaLysAsnLallethr 180
DB 8556 AACGCGCGCGCGCACTCC-----ATCAGCAGCGCGGTGCGGCGAGACTCGCGCTGACC 8609
QY 181 ValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsnHis 200
DB 8610 GTGCGCGCC-----GTGCACTCCACCGACAGC 8636
```



245 SerLysTyrAlaTyrMetGlyThrSerMetAlaThrProIleValAlaGlyAsnVal 264  
1879 -----ATGTCGGAACCTTCGATGGCTACTCCACATGTCAGCGGTGCTGTT 1923  
265 AlaGlnLeuArgGluHisPheValLysAsnArgGlyValThrProLysProSerLeuLeu 284  
1924 GCATCTCTCATAAGCGG---GCAAGGCGGAGGGAATATACTACATCCAGATATAT 1980  
285 LysAlaAlaLeuIleAlaGlyAlaAla-----AspValGlyLeuGly 298  
1981 AGAAGGTTCTTGAGAGCGGTGCAACCTGGCTTGAGGAGATCCATATATCGGAGAG 2040  
299 PheProAsnGlyAsnGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsnValAla 318  
2041 TACACTGAGCTTCACCAAGGTCAATGCTTGTGTAACTTACCAAGTCTCGGAATTCCT 2100  
319 PheValAsnGlnThrSerProLeuSer-----ThrSerGlnLys 331  
2101 AAGGCTATAA---ACGGCACCACCTCTCCCAATTGTTGATCTGCGGAGACAACTCTACA 2158  
332 AlaThrTyrSerPheThrAlaGlnAlaGlyLysProLeuLysIleSerLeuValTrpSer 351  
2159 GCGACT-----TTGCGGAGTACTTGGGTGTGGAGC 2188  
352 AspAlaProGlySerThrAlaSerLeuThrLeuValAsnAspLeuValIle 371  
2189 TTATAAGAGTCTTACGCGAAGCACTCTATACCTG----- 2224  
372 ThrAlaProAsnGlyThr---LysTyrValGlyAsnAsp----- 383  
2225 ACATGTCGAGTGGCACATTAACTAGTACGTAGGGACACCGAGTACAGAACTTTTGAGATCT 2284  
384 --PheThrAlaProTyrAspAsnAsnTrpAspGly-----ArgAsnAsnV 398  
2285 ATGCACTAGCCATGATTAAGCCTTTTCTCAGTGAAGTGAATTTCTAGAGACAATA 2344  
398 alGluAsnValPhe-----IleAsnAlaProGlnSerGlyThrTyr----- 411  
2345 CCGAGTTTGCTTAGGGTGAATATATGATGTAGAGGTCTTTGAGCCAGGTCTCTATGTTG 2404  
412 -----ThrValGluValGlnAlaTyrAsnValP 421  
2405 GAAGGATAATCATTGATGATCCAAACAGCCAGTATTGAAGACGAGATCTTGACACAA 2464  
421 roValSerProGlnThrPheSer 428  
2465 TTGTTATCCGAGAGTTCACT 2487

## RESULT 12

US-10-084-846A-113  
; Sequence 113, Application US/10084846A  
; Publication No. US20040006026A1  
; GENERAL INFORMATION:  
; APPLICANT: WEITNAUER, GABRIELE  
; APPLICANT: MUHLERWEG, AGNES  
; APPLICANT: TREFFER, AXEL  
; APPLICANT: BECHTHOLD, ANDREAS  
; TITLE OF INVENTION: AVILAMYCIN DERIVATIVES  
; FILE REFERENCE: 1974-005  
; CURRENT APPLICATION NUMBER: US/10/084,846A  
; CURRENT FILING DATE: 2003-02-25  
; PRIOR APPLICATION NUMBER: PCT/EP01/09815  
; PRIOR FILING DATE: 2001-08-24  
; PRIOR APPLICATION NUMBER: DE 101 09 166.4  
; PRIOR FILING DATE: 2001-02-25  
; NUMBER OF SEQ ID NOS: 120  
; SOFTWARE: PatentIn Ver. 3.2  
; SEQ ID NO 113  
; LENGTH: 1560  
; TYPE: DNA  
; ORGANISM: Streptomyces viridochromogenes  
; FEATURE:  
; OTHER INFORMATION: avix16 dna: partial sequence of coding strand 1; nucleotide 1

; OTHER INFORMATION: corresponds to nucleotide 56,198 of coding strand 1.  
US-10-084-846A-113

## Alignment Scores:

Pred. No.:	2,83e-22	Length:	1560
Score:	297.00	Matches:	116
Percent Similarity:	39.95%	Conservative:	51
Best Local Similarity:	27.75%	Mismatches:	161
Query Match:	13.19%	Indels:	90
DB:	15	Gaps:	19

US-09-985-689A-7 (1-433) x US-10-084-846A-113 (1-1560)

Qy 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40  
Db 451 GGGCAGGAGTACCGGCTAGCTATCATCACACCGCGTC-----CGC 492  
Qy 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyr---AlaLeuGlyArgThr 59  
Db 493 ATCACCACACGCACTTCGGCGCGCGCTCTACGGCTACGACGCGCATCGACACGAC 552  
Qy 60 AsnAsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn 79  
Db 553 AACACCGCCAGGACGCGCACGCGCACGCGTGGCGGCGACGTCGCCGCGAAC 612  
Qy 80 AlaThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 99  
Db 613 GCC-----TACGGCGTCCCAAGAGCAAGATCGTAGCGTCGCGGTCTCTCAACAAC 666  
Qy 100 GlyGly-----GlyLeuGlyGlyLeuProAlaAsnLeu 110  
Db 667 TCCGCCAGGACACACCGCCAGGTGTCGCCCGGCATCGACTGGTGGTCCCGGAAC--- 723  
Qy 111 GlnThrLeuPheSerGlnAlaTyrSerAlaGlyAlaArgIleHisThrAsnSerTrpGly 130  
Db 724 -----GCCGTCAAGCGCGCGTCCGCAACATGTCCTCCGCG 759  
Qy 131 AlaProValAsnGlyAlaTyrThrThrAspSerArgAsnValAspAspTyrValArgLys 150  
Db 760 GCGCGCGCGACACGCGCTCGACACGCGCGTACGCAAC-----GCCATGCGC 807  
Qy 151 AsnAspMetThrIleLeuPheAlaAlaGlyAsnGlyProGlySerGlyThrIleSer 170  
Db 808 TCCGCGCTACCTTCGCGCGTGGCGCGCAACGAGTCGACCAACGCTCCACGAGGTCA 867  
Qy 171 AlaProGlyThrAlaLysAsnAlaIleThrValGlyAlaThr-----GluAsnLeuArg 188  
Db 868 ---CCGCGACGCGTCACCGAGGCCATCACGCTCGCGCGAGCAGCCAGCTCGACGCCAAG 924  
Qy 189 ProSerPheGlySerTyrAlaAspAsnIleAsnHisValAlaGlnPheSerSerArgGly 208  
Db 925 GCGGCTACTCCAACTACGCTCCCTCTCTC----- 954  
Qy 209 ProThrArgAspGlyArgIleLysProAspValMetAlaProGlyThrTyrIleLeuSer 228  
Db 955 -----GACCTTCGCGCGCGTTCGTCATCATCCTCG 987  
Qy 229 AlaArgSerSerLeuAlaProAspSerSerPheTrpAlaAsnHisAspSerLysTyrAla 248  
Db 988 GCC-----TGAACTCAAGCACTCGCGCGCAAC 1017  
Qy 249 TyrMetGlyGlyThrSerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeuArg 268  
Db 1018 ACATCTCGGTACCTCGATGCGACCGCGCACGTCGGCGGCGCGCGCGCTC----- 1071  
Qy 269 GluHisPheValLysAsnArgGlyValThrProLys-----ProSerLeuLeuLys 285  
Db 1072 ---CACCTCGCGCGCAACCCCTCGCGCACCCCGTCCAGGTGCGCACGCGGTGACGTC 1128  
Qy 286 AlaAlaLeuIleAlaGlyAlaAlaAspValGlyLeuGlyPheProAsnGlyAsnGlnGly 305  
Db 1129 GCGGCCACACCGCGGTGCTGTCACCAACCCCGCACGGGCTCGCCCAACCGGGTCTCTGTAC 1188

QY 306 TrpGlyArgValThrLeuAspLysSerLeuAsnValAlaPheValAsnGluThrSerPro 325  
DB 1189 GTCCGGCGG---CGGCACGACCCCTCCGGCGCCCGGCTTCAGAAC---ACCGGTGAC 1242  
QY 326 LeuSerThrSerGlnLysAlaThrTyrSerPheThrAlaGlnAlaGlyLysProLeuLys 345  
DB 1243 TACACGATACGACCACTCCAGGTGAG---TCCCGGTGACG 1284  
QY 346 IleSerLeuValTrpSerAspAlaProGlySerThrThrAlaSerLeu-----361  
DB 1285 GTCTCCGGCTCCCGCAACGGCCCTCGCGCCCTCGCGCTAGAGGTCCACATCGTCCAC 1344  
QY 362 ThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrLysTyrValGly 381  
DB 1345 ACGTACATCGCGACCTCCAGTCCAGTCCAGTCCAGTCCAGTCCAGTCCAGTCCAGTCC 1404  
QY 382 AsnAspPheThrAlaProTyrAspAsnAsnTrpAspGlyArgAsnAsnValGluAsnVal 401  
DB 1405 AAGTCGTAC-----GGCACCAGCGGCGGAGTTCGGACAACTCAACACACG 1449  
QY 402 Phe---IleAsnAlaProGln-----SerGlyThrTyrThrValGluVal 415  
DB 1450 TACTCGGTGACGCTCTCCGAGCGCGCAACGCGACGCTGGAACTCGCGGTG 1503  
RESULT 13  
US-10-084-846A-1  
; Sequence 1, Application US/10084846A  
; Publication No. US2004006026A1  
; GENERAL INFORMATION:  
; APPLICANT: WEITNAUER, GABRIELE  
; APPLICANT: TREPFER, AXEL  
; APPLICANT: MUELENWEG, AGNES  
; APPLICANT: BECHTOLD, ANDREAS  
; TITLE OF INVENTION: AVILAMYCIN DERIVATIVES  
; FILE REFERENCE: 1974-005  
; CURRENT APPLICATION NUMBER: US/10/084,846A  
; CURRENT FILING DATE: 2003-02-25  
; PRIOR FILING DATE: 2001-08-24  
; PRIOR APPLICATION NUMBER: PCT/EP01/09815  
; PRIOR FILING DATE: 2001-02-25  
; PRIOR APPLICATION NUMBER: DE 101 09 166.4  
; NUMBER OF SEQ ID NOS: 120  
; SOFTWARE: PatentIn Ver. 3.2  
; SEQ ID NO 1  
; LENGTH: 59816  
; TYPE: DNA  
; ORGANISM: Streptomyces viridochromogenes  
US-10-084-846A-1  
Alignment Scores:  
Pred. No.: 5,38e-20 Length: 59816  
Score: 297.00 Matches: 116  
Percent Similarity: 39.95% Conservative: 51  
Best Local Similarity: 27.75% Mismatches: 161  
Query Match: 13.19% Indels: 90  
DB: 15 Gaps: 19  
US-09-985-689A-7 (1-433) x US-10-084-846A-1 (1-59816)  
QY 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40  
DB 56648 GGGCAGGGAGTGACGCGGTACGTATCGACACCGCGCTC-----CGC 56689  
QY 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyr---AlaLeuGlyArgThr 59  
DB 56690 ATCACCACAGGACTTCGGCGCGCGGCTCTACGCTAGCAGCCATCGACACACGAC 56749  
QY 60 AsnAsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn 79  
DB 56750 AACACCGCCAGGACCGCCACGCGCACGCGACGCTGGCGCGGACGCTCGCGGCAAC 56809  
QY 80 AlaThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 99  
DB 56810 GCC-----TACGCGCTGCCCAAGAGGCCCAAGATCGTAGCGTCCGCTGCTGAACAAC 56863  
QY 100 GlyGly-----GlyLeuGlyGlyLeuProAlaAsnLeu 110  
DB 56864 TCCGCGCAGGCACACACCGCCAGGTGTCGCGCATCGACTGGGTGCGCCGGAAC---56920  
QY 111 GlnThrLeuPheSerGlnAlaTyrSerAlaGlyAlaArgIleHisThrAsnSerTrpGly 130  
DB 56921 -----GCCGTCAAGCGCGCGCTCGCCCAACATGCTCCTCGGC 56956  
QY 131 AlaProValAsnGlyAlaTyrThrThrAspSerArgAsnValAspAspTyrValArgLys 150  
DB 56957 GCGCGCGCGCACAGGCCCTCGACACGCGCTGACCAAC-----GCCATGGCC 57004  
QY 151 AsnAspMetThrIleLeuPheAlaAlaGlyAsnGlyProGlySerGlyThrIleSer 170  
DB 57005 TCCGCGCTCACCTTCGCGTGGCGCGGCAACAGTCCGACACGCGCTCCACGAGTCA 57064  
QY 171 AlaProGlyThrAlaLysAsnAlaIleThrValGlyAlaThr-----GluAsnLeuArg 188  
DB 57065 ---CCGCGACGCGTCCAGAGGCCATCAGGTCCGCGGCGGACGACGCTCGGACGCCAAG 57121  
QY 189 ProSerPheGlySerTyrAlaAspAsnIleAsnHisValAlaGlnPheSerSerArgGly 208  
DB 57122 GCGCGCTACTCCAACTACGCTCGCTCCTC-----57151  
QY 209 ProThrArgAspGlyArgIleLysProAspValMetAlaProGlyThrTyrIleLeuSer 228  
DB 57152 -----GACCTCTTCGCGCGCGGTCTGCTCATCATCTCG 57184  
QY 229 AlaArgSerSerLeuAlaProAspSerPheTrpAlaAsnHisAspSerLysTyrAla 248  
DB 57185 GCC-----TGAACCTCAAGCGACTCGGCGACCAAC 57214  
QY 249 TyrMetGlyGlyThrSerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeuArg 268  
DB 57215 ACCATCTCCGCTACGTGATGCGACCCCGCACGCTGGCGGCGCGCGCGCTC-----57268  
QY 269 GluHisPheValLysAsnArgGlyValThrProLys-----ProSerLeuLeuLys 285  
DB 57269 ---CACCTCGCGCGCAACCCCTCGCGCCACCCCGTCCAGGTCCGCGCGCTGACGTC 57325  
QY 286 AlaAlaLeuIleAlaGlyAlaAlaAspValGlyLeuGlyPheProAsnGlyAsnGlnGly 305  
DB 57326 GCGCGCACCCGCGGTCTGTCACCAACCCCGCGCACGCGCTCGCCCAACCGGCTCCTGTAC 57385  
QY 306 TrpGlyArgValThrLeuAspLysSerLeuAsnValAlaPheValAsnGluThrSerPro 325  
DB 57386 GTCCGCGCG---CGGCACGACCACTCCGCGCGCGCTTCGAGAAC---ACCGGTGAC 57439  
QY 326 LeuSerThrSerGlnLysAlaThrTyrSerPheThrAlaGlnAlaGlyLysProLeuLys 345  
DB 57440 TACACGATCAGCGACAACTCCACGCTGAG-----TCCCGGTGACG 57481  
QY 346 IleSerLeuValTrpSerAspAlaProGlySerThrThrAlaSerLeu-----361  
DB 57482 GTCTCCGGGTCTCCCGCAACCGCGCTCGCGCTCGCGTACAGGTCCACATCGTCCAC 57541  
QY 362 ThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrLysTyrValGly 381  
DB 57542 ACGTACATCGCGACCTCCAGGTCCAGTCCAGTCCAGTCCAGTCCAGTCCAGTCCAGTCC 57601  
QY 382 AsnAspPheThrAlaProTyrAspAsnAsnTrpAspGlyArgAsnAsnValGluAsnVal 401  
DB 57602 AAGTCGTAC-----GGCACCAGCGGCGGAGTTCGGACAACTCAACACACG 57646  
QY 402 Phe---IleAsnAlaProGln-----SerGlyThrTyrThrValGluVal 415  
DB 57647 TACTCGGTGACGCTCTCCGAGCGCGCAACGCGACGCTGGAACTCGCGGTG 57700  
RESULT 14  
US-10-084-846A-2/c  
; Sequence 2, Application US/10084846A

```

; Publication No. US20040006026A1
; GENERAL INFORMATION:
; APPLICANT: WEITNAUER, GABRIELE
; APPLICANT: MOHLENWEG, AGNES
; APPLICANT: TREFFER, AXEL
; APPLICANT: BECHTHOLD, ANDREAS
; TITLE OF INVENTION: AVILAMYCIN DERIVATIVES
; FILE REFERENCE: 1974-005
; CURRENT APPLICATION NUMBER: US/10/084,846A
; CURRENT FILING DATE: 2003-02-25
; PRIOR APPLICATION NUMBER: PCT/EP01/09815
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: DE 101 09 166.4
; PRIOR FILING DATE: 2001-02-25
; NUMBER OF SEQ ID NOS: 120
; SOFTWARE: PatentIn Ver. 3.2
; SEQ ID NO 2
; LENGTH: 59816
; TYPE: DNA
; ORGANISM: Streptomyces viridochromogenes
; US-10-084-846A-2

Alignment Scores:
Pred. No.: 5,38e-20 Length: 59816
Score: 297.00 Matches: 116
Percent Similarity: 39.95% Conservative: 51
Best Local Similarity: 27.75% Mismatches: 161
Query Match: 13.19% Indels: 90
DB: 15 Gaps: 19

US-09-985-689A-7 (1-433) x US-10-084-846A-2 (1-59816)

Qy 21 GlycInGlyInleValAlaAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 3169 GGGCAGGAGTACGGGGTACATCGACACCGGGTCC-----CGC 3128
Qy 41 SerMethISgluAlaPheArgGlyLysIleThrAlaLeuTyr---AlaLeuGlyArgThr 59
Db 3127 ATACCACAGGACNTTCGGCGCGCGGGCTCTACGGTACGACCCATCGACACGAC 3068
Qy 60 AsnAsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn 79
Db 3067 AACACCCCGCAGGACGGCCACCGCACCGTGGCGCGCACGGTGGCGCGCAAC 3008
Qy 80 AlaThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 99
Db 3007 GCC-----TACGGCGTGGCCAAAGAGGCCAAGCGTACGGCTCGCGTGTGAACAAC 2954
Qy 100 GlyGly-----GlyLeuGlyGlyLeuProAlaAsnLeu 110
Db 2953 TCGGCGCAGGACACCCGCCAGGTGCTCGCGGCATCGACTGGTGGCGCGAAC--- 2897
Qy 111 GlnThrLeuPheSerGlnAlaTyrSerAlaGlyAlaArgIleHisThrAsnSerTrpGly 130
Db 2896 -----GCCGTCAAAGCCGGCGTCCGCCAACATGTCCCTCCGGC 2861
Qy 131 AlaProValAsnGlyAlaTyrThrThrAspSerArgAsnValAspTyrValArgLys 150
Db 2860 GCGCGCGCGACACGGCCCTCGACACCGCGTACGCAAC-----GCCATGGCC 2813
Qy 151 AsnAspMetThrIleLeuPheAlaAlaGlyAsnGluGlyProGlySerGlyThrIleSer 170
Db 2812 TCCGGCGTACCTTCGGCGTGGCGCGGCAACGAGTCGACCAACGCTCCACGAGGTCA 2753
Qy 171 AlaProGlyThrAlaLysAsnAlaIleThrValGlyAlaThr-----GluAsnLeuArg 188
Db 2752 ---CCCGCAGCGTACCGAGGCCCATCAGGTTCGGCGCGACGACCTCGGACGCCAAG 2696
Qy 189 ProSerPheGlySerTyrAlaAspAsnIleAsnHisValAlaGlnPheSerArgGly 208
Db 2695 GCGGGCTACTCAACTACGGTCCGTCTCTC----- 2666
Qy 209 ProThrArgAspGlyArgIleLysProAspValMetAlaProGlyThrTyrrIleLeuSer 228

```

```

Db 2665 -----GACTCTTCGCGCGCGGTTCGTCATCACCTCG 2633
Qy 229 AlaArgSerSerLeuAlaProAspSerSerPheThrAlaAsnHisAspSerLysTyrAla 248
Db 2632 GCC-----TGGAACTCAACGCGACTCGCGACCAAC 2603
Qy 249 TyrMetGlyGlyThrSerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeuArg 268
Db 2602 ACCATCTCCGTACGTGATGGGACCCCGCACGTGGCGGCGCGCGCTC----- 2549
Qy 269 GluHisPheValLysAsnArgGlyValThrProLys-----ProSerLeuLeuLys 285
Db 2548 ---CACCTCGCGCCCAACCCCTCGGCCACCGTCCAGTTCGCGCACGGCGCTACCTCC 2492
Qy 286 AlaAlaLeuIleAlaGlyAlaAlaAspValGlyLeuGlyPheProAsnGlyAsnGlnGly 305
Db 2491 GCGGCCACCCCGCGGTGTCACCAACCCCGCACGGGTTCGCCCAACCGCTCTCTGTAC 2432
Qy 306 TrpGlyArgValThrLeuAspLysSerLeuAsnValAlaPheValAsnGluThrSerPro 325
Db 2431 GTCGGCGCGG---CGGCACGACACCCCTCCGGCGCGCGCTTCGAGAAC---ACCGGTGAC 2378
Qy 326 LeuSerThrSerGlnLysAlaThrTyrSerPheThrAlaGlnAlaGlyLysProLeuLys 345
Db 2377 TACAGATCAGCGACAACTCCACGTCGAG-----TCCCCCGGTGACG 2336
Qy 346 IleSerLeuValTrpSerAspAlaProGlySerThrThrAlaSerLeu----- 361
Db 2335 GTCTCCGGCGTCTCCGGCACGCGCTCGGCCCTCGCGTAGAGTCCCATCGTCCAC 2276
Qy 362 ThrIleValAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrLysTyrValGly 381
Db 2275 ACGTACATCGCGGACCTCCAGTCCAGTCCAGTCCCGCCCGCGCACGGGTACACGCTC 2216
Qy 382 AsnAspPheThrAlaProTyrAspAsnAsnTrpAspGlyArgAsnAsnValGluAsnVal 401
Db 2215 AAGTCGTAC-----GGCACCGCGCGCAGTTCGCGACAACATCAACACCCACG 2171
Qy 402 Phe---IleAsnAlaProGln-----SerGlyThrTyrThrValGluVal 415
Db 2170 TACTCGGTGAACGCTCTCTCGAGCGCGCCAAACGCGACGTGGAAACTGCGGGTG 2117

RESULT 15
US-09-927-827-29
; Sequence 29, Application US/09927827
; Publication No. US20030036176A1
; GENERAL INFORMATION:
; APPLICANT: Bower, Stanley G.
; APPLICANT: Ramseier, Thomas M.
; TITLE OF INVENTION: Directed Genetic Engineering of Xanthomonas campestris
; FILE REFERENCE: 38-10(15824)B
; CURRENT APPLICATION NUMBER: US/09/927,827
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: US 60/279,493
; PRIOR FILING DATE: 2001-03-28
; NUMBER OF SEQ ID NOS: 69
; SEQ ID NO 29
; LENGTH: 3743
; TYPE: DNA
; ORGANISM: Xanthomonas campestris
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1001)..(2743)
; US-09-927-827-29

Alignment Scores:
Pred. No.: 8.08e-20 Length: 3743
Score: 279.50 Matches: 133
Percent Similarity: 38.00% Conservative: 49
Best Local Similarity: 27.77% Mismatches: 152
Query Match: 12.41% Indels: 145
DB: 10 Gaps: 26

```



```
US-09-985-689A-7 (1-433) x US-09-927-827-29 (1-3743)
QY 21 GlyGlnGlyGlnIleValAlaValAlaAspThrClyLeu-----AspThrGly 36
Db 1502 GGCAGCGCAGCGTGGGGGGTATTGATACCGGCATCACAGTATCGCGACCTCAAC 1561
QY 37 ArgAsnAspSerMetHisGluAlaPheArgGlyLysIleThrAlaLeuValAlaLeu 56
Db 1562 GCCAACATCTCGCGGCTACGACTTCATCAGCGATGCGACACCGCAGCGATGCAAC 1621
QY 57 GlyArgThrAsnAsnAlaAsnAsp-----64
Db 1622 GCGCGTCACACCAACGCGCGCGCGAAGGCGACTGTGTACCGCCCAAGATCGCGCGC 1681
QY 65 -----ProAsnGlyHisGlyThrHisValAlaGlySerValLeuGly 78
Db 1682 GGCATTCCCGCGCGCAGCTCCAGCTGCGCAGCCACCATGTGGCGCGCAGCTCGCGGCA 1741
QY 79 AsnAlaThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuVal-----92
Db 1742 GTGACCAACAACACACCGCGGTAGCGCGCACCGCCTACCGCGCCAAGTCTGATCCGGTG 1801
QY 93 -----PheGlnSerIleMetAsp-----98
Db 1802 CGCGTGTCTCGCAAGTGGGTGGTGTGTCGGATATCGCGGAGCCATCGTCTGGGCC 1861
QY 99 SerGlyGlyLeuGlyGlyLeuProAlaAsnLeuGln-----ThrLeuPheSerGln 116
Db 1862 TCCGGCGCACCGTCCAGCGGCATCCCGGCCAATGCTAACCGCGCGGAGGTGATCAACATG 1921
QY 117 AlaTyrSerAlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyVala 136
Db 1922 TCGTCTGGCGCGCGGTAGTGTCTCGACCACTGACAGACGCC-----ATCAACGGTGGC 1978
QY 137 TyrThrThrAspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeu 156
Db 1979 GTGTGCGCGCGCAC-----ACGGTGGTG 2002
QY 157 PheAlaAlaGlyAsnGluGlyProGly---SerGlyThrIleSerAlaProGlyThrAla 175
Db 2003 GTCGGCGCGCGCAAGATGCGTCCATGTGTCGGT-----TCGCTCGCGGCACTGC 2056
QY 176 LysAsnAlaIleThrValGlyAlaThrGluAsn-----LeuArgProSerPheGlySer 193
Db 2057 GCGAACGTGATTGCGGTGGCGGCCACCACTCGCGCGCGCGAAGCGCAGTATTCCAAC 2116
QY 194 TyrAlaAspAsnIleAsnHisValAlaGlnPheSerSerArgGlyProThrArgAspGly 213
Db 2117 TTCGGCACCGGTATC-----2131
QY 214 ArgIleLysProAspValMetAlaProGlyThrTyrlleLeuSerAlaArgSerSer--- 232
Db 2132 -----GATGTGCGCGCGCGCGTCTCGATCTCTCCAGCTCAACAGCGGC 2179
QY 233 ---LeuAlaProAspSerSerPheTrpAlaAsnHisAspSerLysTyrAlaTyrMetGly 251
Db 2180 ACCACACCGCGGGTAGC-----GCCAGCTATGCTCTCTACAC 2218
QY 252 GlyThrSerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPhe 271
Db 2219 GGCACCTCGATGCGTCCGCGCATGTGGCGCGCGGTGCGCGTGTGTCAGTCG-----2272
QY 272 ValLysAsnArgGlyValThrProLys-----ProSerLeuLeuLys-----AlaAla 287
Db 2273 GTCGCGCGCGCGCTGACCGCGCAGCGGTGGAAACCTTGTGAAAGAACACCGCGCGT 2332
QY 288 LeuIleAlaGlyAlaAlaAspValGlyLeuGlyPhe-----299
Db 2333 GCTTACCGCGCGCGCTGCTCGCGCGGTGCGGTGCGCGCATGCTCAACCGCGATCGCGCG 2392
QY 300 -----ProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAsp 312
```

Search completed: April 5, 2004, 04:52:56  
Job time : 7355.25 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: April 3, 2004, 23:40:05 ; Search time 2274.58 Seconds  
(without alignments)  
5684.703 Million call updates/sec

Title: US-09-985-689A-7  
Perfect score: 2252  
Sequence: 1 NDVARGIVKADVAQNNFGLY.....EVQAYNVFSPQTFSLAIVH 433

Scoring table: BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 27513289 seqs, 14931090276 residues  
Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters: -DEV=xlh  
-Q/cgn2\_1/USPTC.spool/US09985689/runat\_31032004\_161807\_4168/app\_query.fasta\_1.3498  
-DB=EST -QFMT=fastap -SUFFIX=rst -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0  
-UNITG=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pcr -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09985689 @CNC\_1\_1\_7084@runat\_31032004\_161807\_4168 -NCPU=6 -ICPU=3  
-NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THRADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST.\*  
1: em\_estba.\*  
2: em\_esthum.\*  
3: em\_estin.\*  
4: em\_estmu.\*  
5: em\_estov.\*  
6: em\_estpl.\*  
7: em\_estro.\*  
8: em\_htc.\*  
9: gb\_estl.\*  
10: gb\_est2.\*  
11: gb\_htc.\*  
12: gb\_est3.\*  
13: gb\_est4.\*  
14: gb\_est5.\*  
15: em\_estfun.\*  
16: em\_estom.\*  
17: em\_gss\_hum.\*  
18: em\_gss\_inv.\*  
19: em\_gss\_pln.\*  
20: em\_gss\_vrt.\*  
21: em\_gss\_fun.\*  
22: em\_gss\_mam.\*  
23: em\_gss\_mus.\*  
24: em\_gss\_pro.\*  
25: em\_gss\_rod.\*  
26: em\_gss\_phg.\*  
27: em\_gss\_vrl.\*  
28: gb\_gssl.\*

29: gb\_gss2.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	242	10.7	1605	13	BQ622771 CC Contig
2	235.5	10.5	640	12	BQ395336 BJ395336
3	228	10.1	4198	11	AK029048 Mus muscu
4	212	9.4	594	12	BJ393752 BJ393752
5	209.5	9.3	532	29	TA319010P
6	205.5	9.1	771	14	CA320325 UI-M-FW-
C 7	197	8.7	508	28	AO652212 Sheared D
C 8	197	8.7	2141	13	BQ142519 Contig6 M
9	196	8.7	633	12	BJ369190 BJ369190
10	190.5	8.5	601	12	BJ387574 BJ387574
11	189.5	8.4	665	13	BQ770462 UI-M-FIO-
12	186.5	8.3	675	14	CF727824 UI-M-HB-
13	186	8.3	574	29	TA315H10P
14	182	8.1	4662	11	BC060527 T. brucei
15	181	8.0	3091	11	BC011275 Mus muscu
16	180.5	8.0	718	12	BI750157 EG02_1090
17	180.5	8.0	1002	29	CNS0606B
18	179.5	8.0	895	13	BQ216158 AGENCOURT
19	178.5	7.9	641	12	BJ393925 BJ393925
C 20	178	7.9	530	29	CNS010FO
C 21	176.5	7.8	573	14	CA917626 sav42b10.
C 22	175.5	7.8	716	28	BZ893395 HL2_0177
C 23	174	7.7	616	14	CF138007 UI-PR-BNO
C 24	173.5	7.7	2121	28	BZ424995 100023086
25	172	7.6	614	9	AJ273402 AJ273402
26	172	7.6	681	14	CB690041 CEST-54-B
27	171	7.6	449	12	BJ359939 BJ359939
28	170.5	7.6	576	14	CD295943 StrPu691.
29	170.5	7.6	594	14	CF846026 psHB0362A
30	170.5	7.6	1572	11	AY107161 Zea mays
31	170	7.5	545	13	BU575479 TGEStzY08
32	169	7.5	712	14	CD311344 StrPu691.
33	169	7.5	930	13	BJ390734 BX390734
C 34	168	7.5	564	28	AO651427 Sheared D
35	167	7.4	650	9	AJ274038 AJ274038
36	166	7.4	937	12	BQ246418 602360428
37	165.5	7.3	688	10	BF636040 NF070C09D
38	165.5	7.3	887	13	BQ879057 AGENCOURT
39	165	7.3	601	12	BJ365857 BJ365857
40	165	7.3	732	13	BU028438 QHH14A24.
41	165	7.3	794	14	CF737198 UI-M-HD-
42	164.5	7.3	989	29	CC723222 OGLAB20TV
C 43	164	7.3	655	28	BZ349201 hg88c10.g
44	163	7.2	786	13	BU709126 UI-M-EW-
45	163	7.2	847	14	CF726182 UI-M-GZ-

ALIGNMENTS

RESULT 1  
BQ622771 BQ622771 1605 bp mRNA linear EST 01-JUL-2002  
LOCUS CC Contig67 Conidiobolus cornatus ARSEP 512 Conidiobolus coronatus  
DEFINITION cdna, mRNA sequence.  
ACCESSION BQ622771.1 GI:21649940  
VERSION BQ622771  
KEYWORDS EST  
SOURCE Conidiobolus coronatus (Delacroixia coronata)  
ORGANISM Conidiobolus coronatus  
Eukaryota; Fungi; Zygomycota; Zygomycetes; Entomophthorales;  
Ancylistaceae; Conidiobolus.  
REFERENCE 1 (bases 1 to 1605)

Qy	178	AlaIleThrValGlyAlaThrGlu-----AsnLeuArgProSerPheGlySerTyrAla	195
Db	1258	GCAATACCGTTGGTCCTACTGATGAATAAAAGGCCTCATCTCTTAACCTCCGT	1317
Qy	196	AspAsnIleAsnHisValAlaGlnPheSerSerArgGlyProThrArgSpGlyArgIle	215
Db	1318	AGCTGTGTC-----	1326
Qy	216	LysProAspValMetAlaProGlyThrTyrlLseuSerAlaargSerSerLeuAlaPro	235
Db	1327	-----GATATCTTAGTCTCTGGTGCAACATTCTCTCCACC-----	1362
Qy	236	AspSerSerPheTrpAlaAsnHisAspSerLysTyrAlaIatYrMetGlyGlyThrSerMet	255
Db	1363	-----TGGAAGGGATCTAACACGCCGCCAACCAACCATCTCTGGTACCTCTCATG	1410
Qy	256	AlaThrProIleValAlaGlyAsnValAlaGlnLeuArggGliuHisPheVallysAsnArg	275
Db	1411	GCTTGGCCTCACATTGCTGTTTTAGTGTACTGTCTTCTTCTCAAGTACTGCTCGCCGG	1470
Qy	276	GlyValThrProLysProSerLeuLeuLysAla	286
Db	1471	TGGAGTTTGATGCTTAAG-----CTCAAGGTT	1495

Db	1471	TCGAGTTGATGCTAAG	-----CTCAAGGCT	1495
RESULT 2				
BJ395336				
LOCUS	BJ395336	640 bp	mRNA	linear
DEFINITION	BJ395336	Dictyostellium discoideum	CDNA library, SF	Dictyostellium
		discoideum	CDNA clone dda3816 5'	mRNA sequence.

RECEIVED  
03 03 00  
VERSION BJ395336.1  
GI:19306422

SOURCE	Dictyostelium discoideum
ORGANISM	Dictyostelium discoideum
REFERENCE	Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium. 1 (bases 1 to 640)

REFERENCE  
1 (bases 1 to 540)  
AUTHORS  
TITLE  
JOURNAL  
Unpublished (2002)

**AUTHORS** Urushinara, H., Tanaka, I., Konaka, I. and Shimizu, I.  
**TITLE** Full length cDNA of Dictyostelium discoideum at the slug stage  
**JOURNAL** Unpublished (2002)  
**COMMENT** Contact: Tadasu Shin-i

Unpublished (2002)  
Contact: Tadasu Shin-i  
Center For Genetic Resource Information  
National Institute of Genetics  
1111 Yata, Mishima, Shizuoka 411-8540, Japan

Center For Genetic Resource Information  
National Institute of Genetics  
1111 Yata, Mishima, Shizuoka 411-8540, Japan  
Tel: 81-559-81-6856  
Fax: 81-559-81-6855

Ill. Yata, Mienima,  
Saitzuka 411-8540, Japan  
Tel: 81-559-81-6856  
Fax: 81-559-81-6855  
Email: tsunig@nes.nig.ac.jp.

Fax: 81-559-81-6855  
 Email: [ethin@genes.nig.ac.jp](mailto:ethin@genes.nig.ac.jp)  
 Location/Qualifiers  
 1. 640  
 FEATURES  
 SOURCE  
 1. 640

```

FEATURES
  source
    Location/Qualifiers
      1..640
        /organism="Dictyostelium discoideum"
        /mol type="mrna"

```

```

/organism="Dictyostellium discoideum"
/mol_type="mRNA"
/strain="AX4"
/db_xref="taxon:44689"

```

```

/strain="Ax4"
/db_xref="taxon:44689"
/clone="dda38b1c"
/sex="mat A"

```

```

/clone="dds38b16"
/sex="mat A"
/dev stage="Slug stage"
/clone lib="Dictyostelium discoideum cDNA library, sf"

```

```

ORIGIN
/clone_lib="Dictyostelium discoideum cDNA library, SF"
/dev_stage="slug stage"

```

ORIGIN

Alignment Scores:	4,23e-13	Length:	640
Pred. No.:			

Alignment Scores:	4.23e-13	Length:	640
Pred. No.:	235,50	Matches:	69
Score:	49.77%	Conservative:	37
Percent Similarity:			

Score:	235.50	Matches:	69
Percent Similarity:	49.77%	Conservative:	37
Best Local Similarity:	32.39%	Mismatches:	68
Overall Match:	10.46%	Indels:	39

Best Local Similarity:	32.39%	Mismatches:	68
Query Match:	10.46%	Indels:	39
DB:	12	Gaps:	10
			12

DB:	12	Gaps:	10
US-09-985-689A-7 (1-433) x BJ395336 (1-640)			

US-09-985-689A-7 (1-433) x BJ395336 (1-640)  
QY 19 LeuTyrGlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArg --- 37

Qy 19 LeuTyrGlyVGLnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArg--- 37  
Db 22 TTAAGAGGTAAAGGTCAGATATTCAGTATTGCTGATCTGATTGTTAGATGTTAGCCATTGT 81

Db 22 TTTAAGAGGTAAAGCTCAGATATTCAGTATTGCTGATCTGGTTTAGAIGTAGCCATTGT 81

QY 38 -----AsnAspSer-----SerMetHisGluAlaPheArgGly 48  
Db 82 TTTCTTTTCAGATTCAAGTATCCAAATACCAATTAATCAAGTCAATCAAAATCATAGAAA 141  
QY 49 LysIleThrAlaLeuTyAlaLeuGlyArgThrAsnAsnAlaAsnAspPro-----Asn 66  
Db 142 GTTGTAACATTATAT-----ACTTACCATGACAATGAAGATTATGTAAT 186  
QY 67 GlyHisGlyThrHisValAlaGlySerValLeuGlyAsn----- 79  
Db 187 GTCATGGTACACATGTTTGTGGCTCTGCAGCAGTACTCCAGAGGATCTTCATGGGCT 246  
QY 80 ----AlaThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAsp 98  
Db 247 ATTTCATCATTTAGTGGTCTTGCAGTACGACGAAGATTGCAATTTATGAT---CTTCA 303  
QY 99 SerGlyGlyGlyLeuGlyLeuProAlaAsnLeuGlnThrLeuPheSerGlnAlaTy 118  
Db 304 TCTGGAAGTCTCAACCAACACACCACCGAAGATTACAGTCAATGTACAAACCATATAT 363  
QY 119 SerAlaGlyAlaArgIleHisThrAsnSerTrpGlyAla-----ProValAsn 134  
Db 364 GATGCAGGTGCAAGACATACATGTTGATCTTGGGTTCTGTATCTTTGCAAGGTTATAT 423  
QY 135 GlyAlaTyThrThrAspSerArgAsnValAspAspTyValArgLys---AsnAspMet 153  
Db 424 GFGGTTATCCGATGATGCTGGTGTATGATGCAATCTCTATGAGTACCAAGATTC 483  
QY 154 ThrIleLeuPheAlaAlaGlyAsnGlyProGlySerGlyThrIleSerAlaProGly 173  
Db 484 TCTATACTAAGAGCTGCTGTAAT---AAGAGCATTTTGCATCTTTTAGCTCAAGCA 540  
QY 174 ThrAlaLysAlaIleThrValGlyAlaThrGluAsnLeuArgProSerPheGlySer 193  
Db 541 ACAGCTAAAAATGCAATTACATGTTGGTGTGACAAACAGCTCATGTAATTTATGTCGA 600  
QY 194 -----TyrAlaAspAsnIleAsn 199  
Db 601 GATGATTCGATATATGATTTCTCAGATAATGCTAAT 639  
RESULT 3  
AK029048  
LOCUS  
DEFINITION  
AK029048 4198 bp mRNA linear HTC 18-SEP-2003  
Mus musculus 10 days neonate skin cDNA, RIKEN full-length enriched  
library, clone:473248M11 product:membrane-bound transcription  
factor protease, site 1, full insert sequence.  
AK029048  
VERSION  
AK029048.1 GI:26325017  
KEYWORDS  
Mus musculus (house mouse)  
SOURCE  
Mus musculus  
ORGANISM  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE  
1 Carninci, P. and Hayashizaki, Y.  
TITLE  
High-efficiency full-length cDNA cloning  
JOURNAL  
Meth. Enzymol. 303, 19-44 (1999)  
MEDLINE  
99279253  
PUBMED  
10349636  
REFERENCE  
2  
AUTHORS  
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,  
Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
TITLE  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new genes  
JOURNAL  
Genome Res. 10 (10), 1617-1630 (2000)  
MEDLINE  
20499374  
PUBMED  
11042159  
REFERENCE  
3  
AUTHORS  
Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,  
Kono, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M.,  
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,  
Yanamoto, S., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,  
Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M.,

Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,  
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.  
RIKEN integrated sequence analysis (RISA) system--384-format  
sequencing pipeline with 384 multicapillary sequencer  
Genome Res. 10 (11), 1757-1771 (2000)  
20530913  
11078861  
4  
The RIKEN Genome Exploration Research Group Phase II Team and the  
FANTOM Consortium.  
Functional annotation of a full-length mouse cDNA collection  
Nature 409, 685-690 (2001)  
5  
The FANTOM Consortium and the RIKEN Genome Exploration Research  
Group Phase I & II Team.  
Analysis of the mouse transcriptome based on functional annotation  
of 60,770 full-length cDNAs  
Nature 420, 563-573 (2002)  
6  
(bases 1 to 4198)  
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,  
Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,  
Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,  
Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,  
Kato, H., Kawai, J., Kojima, Y., Kondo, S., Kono, H., Kouda, M.,  
Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,  
Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N.,  
Okazaki, Y., Saiki, D., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,  
Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,  
Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,  
Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,  
Muramatsu, M. and Hayashizaki, Y.  
Direct Submission  
Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of  
Physical and Chemical Research (RIKEN), Laboratory for Genome  
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),  
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,  
Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,  
URL: http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,  
Fax:81-45-503-9216)  
cDNA library was prepared and sequenced in Mouse Genome  
Encyclopedia Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in RIKEN  
Division of Experimental Animal Research in Riken contributed to  
prepare mouse tissues.  
Please visit our web site for further details.  
URL: http://genome.gsc.riken.go.jp/  
URL: http://fantom.gsc.riken.go.jp/  
FEATURES  
source  
Location/Qualifiers  
1..4198  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="FANTOM.DB:473248M11"  
/db\_xref="MGI:2351080"  
/db\_xref="taxon:10090"  
/clone="473248M11"  
/tissue\_type="skin"  
/clone\_lib="RIKEN full-length enriched mouse cDNA library"  
/dev\_stage="10 days neonate"  
417..3575  
/note="unnamed protein product; membrane-bound  
transcription factor protease, site 1 (MGD|MGI:1927235,  
GB|NM\_019709, evidence: BLASTN, 99%, match=3782)  
putative"  
/codon\_start=1  
/protein\_id="BAC36263.1"  
/db\_xref="GI:26325018"  
/translation="MKLVSTWLVLLVCGRLHGLDRLGTRALEKAPCPSCSHLTK  
VERSTVVEYIVAFYGTAKARNSFISALKSEVENWR11PRNNPSDYSDPDE  
VIQKEQKAGLLIEDHPNKRVTPOKVPKSLKPAESNPVPCNTRWQKQSSR  
TLKASLUSGFWHATGRHSRLRLRAIPQVATQGLQDVLMQNGYTCANRVAVFD  
PGLSEKHPFKVQWRTNRTNLTDLGLGTFVAGVISMRCQGPAPDAELHIFR  
VFTNNQVSYTSWFLDAFNAYIALKKMDVLNLISIGGPFDMDFPVDKVMELTANNVIMVS

ATGNDGPIYGTLLNNPADQMDVTVGVGIDFEDNRIAPFSSRGMTTWELPGGYGRVKPDPIY  
 TYGAVRGSGVGGKRCALSGTSVAPAGAVTLTVSTVQKEELVNPASVWQALIASA  
 RLALPGVFEQGGKGLDLARAYQLISYQASLSPSYITDLTCYMPMFYQQALITYG  
 GMPTIVNVTILANGWGTGRIYVQKPEWRYLPQNGDNIYEVAFSYSVLAPWFSYGLAIS  
 SVTKKASMEGTAQGHIMTIVASPAETLHSGAEHTSTVKLPKVKVITPTPRSKRVL  
 MQYHNLYRVPYGVPEEDNIRMKNDPLDNGKDHVHTNFRDMYOHLSRMSGYFVRLVGLAPF  
 TCFDATQYGTGLLVLDSEERYFPEETAALBRDVNGLSLVIFSDWNTYSVMRKAKSYDE  
 NTRQWMPPTGTLLNPNELLUSVMNMFSDGLYEGEFGFVLANHDNTYASGCSIAKFFE  
 DGVYITVTLQDQGLYEQKATAVENVILGHYQIPSPEGGRIYVLYGSDNCLUDDSHRQ  
 KQCFWLLDALLOYTSYGVTFPSLSHSGNRQPPSGAGLAPPPERMEGNHLRYSKVLEA  
 HLGDPKRPPLPACFHLWSAKPOPLNATPSNWKQKLLSIDLKVLPNPNRNPQV  
 RLSPGCEGAWDIPGGIMPRGNQEVGTQIPVFAFLGAMVALAFFVYVQISAKKSRRPKR  
 RAPRAKRPQLAQQAHPARTFSV"

polyA\_signal  
 4180..4185  
 /note="putative"  
 4198  
 /note="putative"

## ORIGIN

Alignment Scores:					
Pred. No.:	5_91e-11	Length:	4198		
Score:	228.00	Matches:	119		
Percent Similarity:	38.46%	Conservative:	51		
Best Local Similarity:	26.92%	Mismatches:	166		
Query Match:	10.12%	Indels:	106		
DB:	11	Gaps:	23		
US-09-385-689A-7 (1-433) x AK029048 (1-4198)					
Qy	8	VallvYAlaAspValAlaGlnAsnAsnPheGlyLeuTyrglyGlnGlyGlnIleValala	27		
		: :::			
Dbb	1005	CTCAGGCACATGTGCTGTGG---CAGATCGGGATACACAGTCTCTAAATGTCAGAGTTGCT	1061		
Qy	28	ValalaaSpThrGlyIleuAspThrGlyArgAsnAspSerMetHisGlualaPheArg	47		
Dbb	1062	GTATTTTGATACTGGGCTC-----AGTCAGAAGCATCCGCATTTTTAAG	1103		
		:   ::			

Qy	48	GlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsnAsnAlaLen-----	63
Db	1104	AAT-----GTCAAGCAGAGAACCAACTCGACCAATGACGGACCCCTG	1145
Qy	64	--AspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnAlaThrAsn	82
Db	1146	GATGATGGGTAGGCCATGGCACATTCTGTCAGGTCTGATTGCCACCATGCGGAGTGC	1205
Qy	83	LysGlyMetAlaProGlnAlaAsnLeu---ValPheGlnSerIleMetAspSerGlyGly	101
Db	1206	CARGGATTGTCTCCAGATGCAGAGCTGCACATCTTCAGGTCTTTACCAACAAT-----	1259
Qy	102	GlyLeuGlyGlyLeuProAlaAsnLeuGlnThrLeuPheSerGlnAlaTyrSerAlaGly	121
Db	1260	-----CAGGTGTCTTACACATCTTGGTTCTGGATGCGCTTCAACTATGCC	1304
Qy	122	AlaArgIleHisThrAsnSerTyrGlyAlaProValAsnGlyAlaTyrThrAspSer	141
Db	1305	ATCCTAAGAAGATGGACGTCTCAACCTTAGCATCGTGGCCCGACTTCATGGATCAT	1364
Qy	142	ArgAsnValAspAspTyr-----ValArgLysAsnAspMetThrIleLeuPheAlaAla	159
Db	1365	CCGTTTGTGCAAGGTGTGGGAATTAAACAGCTAACCAATGAATTATGGTTTCTGCTATT	1424
Qy	160	GlyAsnGluGlyProGlySerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle	179
Db	1425	GGCAATGATGACCTCTCTATGGCACTCTGAAATACCCCTGCTGATCAGATGGATGCTGATT	1484
Qy	180	ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn	199
Db	1485	GGAGTGGGTGCATTGAC-----TTTGAAGATAACATC---	1517
Qy	200	HisValAlaGlnPheSerSerArgGly-----ProThrArgAspGly	213
Db	1518	-----GTCGCTCTTTCTTCAGGGGAATGACTACCTCGGGAATACCAAGGAGGTATGGT	1571

Qy	214	ArgilleysProaspValMetalAlaProGlyThrTyrlleuSerAlaargSerLeu	233
Dd	1572	CGTGAAGCTTCACATTGC-----ACCTATGGTGCTCGAGTGC GG GTTCGGT	1522
Qy	234	AlaProAspSerSerPheTrpAlaAenHisAspSerLysTyzAlaTyzMtGlyGlyThr	283
Dd	1623	GTCAAAAGGGCGCTGC-----CGTGCACTCTTCAGGACC	1655
Qy	254	SerMetAlathPrOtleValAlaglyAenValadlnLeuArgGluHISpheVallys	273
Dd	1656	AGTGTCCCTTCCCAGTGGTCTGGGCCCTCACCTTTAGTAGCAGCAGTACAGAAG	1715
Qy	274	AsnArgglyValThrProLysProSerLysLeuLeuAlalaLeuleAlaGlyAlaAa	293
Dd	1716	CGGAGCTGGTG-----AATCTGCCAGTGGTAAGCAAGCTTTGATAGCGTCAGGCCGG	1769
Qy	294	AspValglyLeuGlyPheProAenGlyAse-----GlnGlyTrpGlyArGValThr	310
Dd	1770	AGACTT-----CTGGGTCAACATGTTGAGCCAAGTTCATGCCAAGTTGGAT	1817
Qy	311	LeuAsphySserLeuAenValAlaPheValAsnGluThrSerProIeuserThrsrGln	330
Dd	1818	CTGTGGAGCTTATCAGATC-----CTCAGC-----	1844
Qy	331	LysAlathrTyxSerPheThrAlaGlnAglay-----LysProLeuLyslleSerLeuVal	349
Dd	1845	-----AGCTATAAACCGCAGCGCAAGCTCGAGTCTAGCTACATCGACCTGACT	1892
Qy	350	-----TrpSerAspAlaProGlySerThrThralaserLeuThr	362
Dd	1893	GAGTGTCCCTACATGTGGCCCTACTGTCCAGCCTACTACTATGAGGAAATGCCAACCA	1952
Qy	363	LeuValasnAspLeuAspLeuValileThralaproAenglyThryltyrvalglyYasn	382
Dd	1953	ATCGTTAAT-----GTCCCATCTCTCAATGGCCTGGCGCTCACAGGAAGA	1997
Qy	383	AspPheThralaProTyraSpAsnaStrp-----AspGlyArGAsnaSn	397
Dd	1998	ATTGTGATAGCCT-----GAGTGGCGACCCCTATTACCACAGAATGGAGACAAC	2048
Qy	398	ValGluasnValPhe-----IleAsnAlaProGlnSerGlyThrTyThrVal	413
Dd	2049	ATTGAGTGGCCCTTCTCTACTCTCAGTGTGTGGCCCTGGTCAGGTACCTTGCCATC	2108
Qy	414	GluVal 415	
Dd	2109	TCCATT 2114	
RESULT 4			
LOCUS	BJ393752		
DEFINITION	BJ393752 Dictyostelium discoideum cDNA library, SF Dictyostellium	594 bp mRNA linear EST 08-MAR-2002	
ACCESSION	Dictyostelium cDNA clone dds32b16 5', mRNA sequence.		
VERSION	BJ393752.1 GI:19304838		
KEYWORDS	EST.		
SOURCE	Dictyostelium discoideum		
ORGANISM	Dictyostelium discoideum		
REFERENCE	Eukaryota; Mycetozoa; Dictyosteliida; Dictyostellium.		
AUTHORS	1 (bases 1 to 594)		
TITLE	Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.		
JOURNAL	Full length cDNA of Dictyostelium discoideum at the slug stage		
COMMENT	Unpublished (2002) Contact: Tadashi Shin-i Center For Genetic Resource Information National Institute of Genetics 1111 Yata, Mishima, Shizuoka 411-8540, Japan Tel.: 81-559-81-6856 Fax: 81-559-81-6855 Email: tshini@genes.nig.ac.jp. Location/Qualifiers 1..594 /organism="Dictyostelium discoideum"		
FEATURES			
source			

FEATURES  
SOURCE

```
1. .594
/organism="Dictyostelium discoideum"
```

/mol\_type="mRNA"  
/db\_xref="taxon:44689"  
/clone="dds32b16"  
/sex="mat A"  
/dev\_stage="Slug stage"  
/clone\_lib="Dictyostelium discoideum cDNA library, SF"

## ORIGIN

## Alignment Scores:

Pred. No.: 8,61e-11 Length: 594  
Score: 212.00 Matches: 54  
Percent Similarity: 50.2% Conservative: 39  
Best Local Similarity: 29.1% Mismatches: 62  
Query Match: 9.41% Indels: 30  
DB: 12 Gaps: 8

US-09-985-689A-7 (1-433) x BU393752 (1-594)

Qy 49 LyslleThrAlaLeuTy:AlaLeuGlyArgThrAsnAsnAlaAsnAspProAsnGlyHis 68  
Db 21 AAAGTGTGAACCTATATTACCATCAACACGACGAGTAGTGAATAAGTGGTGCAC 80  
Qy 69 GlyThrHisValAlaGlySerValLeuGlyAsn-----Ala 80  
Db 81 GGTACACATATTGTGGTCTGCAGCAGGTACTCCAGAGGATTCTTCAGTTAATATTCA 140  
Qy 81 ThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSerGly 100  
Db 141 TCATTAGTGGTCTTGTCACTGATGCAAGATTGCATTC-----TTTGATTGGCA 191  
Qy 101 GlyGlyLeuGlyGlyLeu-----ProAlaAsnLeuGlnThrLeuPheSerGlnAlaTyr 118  
Db 192 AGTGTTTCATCAAGTTTGACACCTCCATCGGATTTGAACAATTATATCAACCAATTAT 251  
Qy 119 SerAlaGlyAlaArgIleHisThrAsnSerTyrGlyAla-----ProValAsn 134  
Db 252 GACGACGGTGCAGAGTGTGATTTCTGGGTCTGTATCATAGTACAGGGGTATACA 311  
Qy 135 GlyAlaTyrThrThrAspSerArgAsnValAspAspTyrValArgLysAsn---AspMet 153  
Db 312 GGTAGTATTCATCAGACACTGCTTCAATTGATGATTTCTTTCATCATCCAGATTTC 371  
Qy 154 ThrIleLeuPheAlaAlaGlyAsnGluGlyProGlySerGlyThrIleSerAlaProGly 173  
Db 372 ATCATTTCTTAGAGTCTGCTGTACACGACGCAATACCTATCATCTACTCATCT 428  
Qy 174 ThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsnLeuArg-----188  
Db 429 ACTGCAAGAATGTTATTACCGTTGGTGTCTCATCAACAATTCATGAAAATTATTAACT 488  
Qy 189 -----ProSerPheGlySerTyrAlaAspAsnIleAsn-----HisValAla 202  
Db 489 GATGGTCCAAATTATATAAATTATCAATCATCTGTGCTGATATAAATCAAGAGTTAATATGT 548  
Qy 203 GlnPheSerSerArg 207  
Db 549 GATTTCGATAGCAGA 563

## RESULT 5

## TA319G10P

## LOCUS

## DEFINITION

T. brucei sheared genomic DNA clone 319g10, forward sequence,  
genomic survey sequence.

## ACCESSION

## VERSION

## AL492464

## KEYWORDS

## SOURCE

## ORGANISM

Trypanosoma brucei  
Trypanosoma brucei  
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;  
Trypanosoma.

## REFERENCE

## AUTHORS

1 (bases 1 to 532)  
Hall,N., Bowman,S., Lennard,N.J., Doggett,J., Atkin,R.,

Chillingworth,C., Ormond,D., Harris,B., El-Sayed,N., Hou,L.,  
Melville,S.E., Rajandream,M.A. and Barrell,B.G.

## TITLE

## JOURNAL

Direct Submission  
Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing  
project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton,  
Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and  
nh@sanger.ac.uk

## COMMENT

Constructed at the Institute for Genomic Research (TIGR),  
Rockville, MD. Genomic DNA isolated from a cloned population of  
Trypanosoma brucei (TREU927/4 CUTat 10.1) was mechanically sheared  
to give a tight size distribution (4 kb). The v + i method used for the library construction is  
described in detail in Smith, H. and Venter, J.C. (Making small  
insert libraries for whole genome shotgun sequencing projects. In  
Genome Sequencing: A Practical Approach, eds. M. Vaudin and B.  
Barrell, Oxford University Press, 1999).

Email: nelsayed@tigr.org  
Details of T. brucei sequencing at the Sanger Centre are available  
at http://www.sanger.ac.uk/Projects/T\_brucei/.

## FEATURES

## source

1..532  
/organism="Trypanosoma brucei"  
/mol\_type="genomic DNA"  
/strain="TREU927"  
/db\_xref="taxon:5691"  
/clone="319g10"

## ORIGIN

Alignment Scores:  
Pred. No.: 1,27e-10 Length: 532  
Score: 209.50 Matches: 61  
Percent Similarity: 47.06% Conservative: 27  
Best Local Similarity: 32.62% Mismatches: 66  
Query Match: 9.30% Indels: 33  
DB: 7 Gaps: 7

US-09-985-689A-7 (1-433) x TA319G10P (1-532)

Qy 92 ValPheGlnSerIleMetAspSerGlyGlyGlyLeuGly-----104  
Db 2 GTAATGCCAAGTAAATGGTGTCCGCCAGGGCGGAAGATTTCCTCAGGGGTGGGTGCC 61  
Qy 105 -----GlyLeuProAlaAsnLeuGlnThrLeuPheSerGlnAlaTyr 118  
Db 62 CATCCAGTCAGAGCTGTCTCTCCCCACGACGTTACTCAAAATTATTCGTCCTGATAT 121  
Qy 119 SerAlaGlyAlaArgIleHisThrAsnSerTyrGlyAlaProValAsnGlyAlaTyr 138  
Db 122 GGCCTGGAGCCCGGTGTCTCTCAAACTCGTGGGGTTTGTCTCCCTCCGAGTATTC 181  
Qy 139 ThrAspSerArgAsnValAspAspTyrValArgLys---AsnAspMetThrIleLeuPhe 157  
Db 182 GCTGTGAAAGGATATGGATGGATTTGGCGAGTAGTATGACGATGGCTACTTATCTTC 241  
Qy 158 AlaAlaGlyAsn-----GluGlyProGlySerGlyThrIleSerAlaProGlyThr 174  
Db 242 TCCACTGCACACGATATCCAAAGATGGC-----CTAATGACTCCGTGTCGT 286  
Qy 175 AlaLysAsnAlaIleThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyr 194  
Db 287 GGTAAAGACCGTATGTCGCGGGGTGCACACAAACCGTG-----TTTGACGCTTCG 337  
Qy 195 AlaAspAsnIleAsnHisValAlaGlnPheSerSerArgGlyProThrArgAspGlyArg 214  
Db 338 AAAGAC-----ATTGTTCTCTTCGTTCTTCGTCATGTCCTCCACATACGCGTAGG 388  
Qy 215 IleLysProAspValMetAlaProGlyThrTyrIleLeuSerAlaArgSerSerLeuAla 234  
Db 389 ATGAACCCGATCTTGTGCGTCCCGGGGAAGAGGTGCTGCTCTTCTTCTCTGGCAA 448  
Qy 235 ProAspSerSerPheTrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSer 254  
Db 449 GCATCAGCTAAA-----CAATGTAAAGTGGTGGCCCAAGCGGNGTTCATCG 493







```

Db 906 CGAAGCTACGGTGTTCGCAAAATCCAGCTCTTGGTGTCAAGGTCTTGTATGACCG 847
Qy 99 SerGlyGlyLeuGlyLeuProAlaLeuThrLeuPheSerGlnAlaTyr 118
Db 846 GGCAGTGGTTCCTACTCCGGTATCATCAGTGGCATGGACTTGTGCGCAGGACTCCAAG 787
Qy 119 Ser-----AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGly 135
Db 786 AGTCGTAACTGCCCAATGGCCACATTCCTCCATGAGTCTGGCA-----GGT 739
Qy 136 AlaTyrThrAspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIle 155
Db 738 GGCTACTCGGCTCGCTCAACAGCGGTCCGCTGTTGGTCAGGTCGTGGTCTTCCCT 679
Qy 156 LeuPheAlaAlaGlyAsnGluGlyProGlySerGlyThrIleSerAlaProGlyThrAla 175
Db 678 GCCGTCGCGCTGGCAAGTAAACCGGATGCCCAAAACACCTCT-----CCGCTTCGAG 622
Qy 176 LysAsnAlaIleThrValGlyAlaThr-----GluAsnLeuArgProSerPheGlySer 193
Db 621 CCTACTGCTGCTGACCTGTTGGTCCACCTGCGTCAGATGACAGCGCATCTACCTTTTCCAAC 562
Qy 194 TyrAlaAspAsnIleAsnHisValAlaGlnPheSerSerArgGlyProThrArgAspGly 213
Db 561 TAC-----TGATCAAT-----GGCGGCCAACACCATCTCTGGCACC 475
Qy 214 ArgIleLysProAspValMetAlaProGlyThrTyrIleLeuSerAlaArgSerSerLeu 233
Db 555 AGAGTT---GTGATATCTTCGCTCGTGTACCGGCATCTCTCCACC-----511
Qy 234 AlaProAspSerSerPheTrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThr 253
Db 510 -----TGATCAAT-----GGCGGCCAACACCATCTCTGGCACC 475
Qy 254 SerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLys 273
Db 474 TCCATGGCTACTCCCATATTCGTGCT-----CTTGCTGCTACTCTCAGTGCT 427
Qy 274 AsnArgGlyValThrProLysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAla 293
Db 426 CTCAGCGGACAGACTAGCCCTCGCTCTTTGCCAGAGATCCAGGACATCTTACCACAG 367
Qy 294 AspValGlyLeuGlyPheProAsnGly 302
Db 366 AACGTGATCCGCAATGTGCGCGCTGCG 340

```

```

RESULT 9
BJ369190 633 bp mRNA linear EST 08-MAR-2002
LOCUS
DEFINITION
Dictyostelium discoideum cDNA library, CF Dictyostelium
ACCESSION
Dictyostelium discoideum cDNA clone ddc49116 S', mRNA sequence.
VERSION
Dictyostelium discoideum
SOURCE
Dictyostelium discoideum
ORGANISM
Dictyostelium discoideum
REFERENCE
1 (bases 1 to 633)
AUTHORS
Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
TITLE
Full length cDNA of Dictyostelium discoideum at the culmination
stage
JOURNAL
Unpublished (2002)
COMMENT
Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.
Location/Qualifiers
1 . 633
/organism="Dictyostelium discoideum"

```

```

FEATURES
Source

```

```

/mol_type="mRNA"
/strain="AX4"
/db_xref="taxon:44689"
/ciome="ddc49116"
/sex="mat A"
/dev stage="Culmination stage"
/clone_lib="Dictyostelium discoideum cDNA library, CF"

```

## ORIGIN

```

Alignment Scores:
Pred. No.: 3.9e-09 Length: 633
Score: 196.00 Matches: 58
Percent Similarity: 41.55% Conservative: 28
Best Local Similarity: 28.02% Mismatches: 49
Query Match: 8.70% Indels: 72
DB: 12 Gaps: 6

US-09-985-689A-7 (1-433) x BJ369190 (1-633)

Qy 135 GlyAlaTyrThrAspSerArgAsnValAspAspTyrValArgLys---AsnAspMet 153
Db 6 GGTGTTATTCGATGATGCTGGTGTGTTGATGATGCTCTATGAGTACCCAGAAATTC 65
Qy 154 ThrIleLeuPheAlaAlaGlyAsnGluGlyProGlySerGlyThrIleSerAlaProGly 173
Db 66 TCTATCTAAGAGCTGCTGCTAAT---AACGACTATTTCATCTTATTATGCTCAAGCA 122
Qy 174 ThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsnLeuArgProSerPheGlySer 193
Db 123 ACAGCTAAAAATGCAATTACAGTTGGTGTGCTGAGCAACAGCTCATGTAAATATTGTGTC 182
Qy 194 -----TyrAlaAspAsnIle-----198
Db 193 GATGATTTGGNAATATTATGATTCTCAGATATGCTAATTTCAAAGACCATGTTTATTC 242
Qy 198 -----198
Db 243 GATAAGAAATGATTGTAATTATACGACGCTAAATGTTGTCAGAGGTTTCAAATGTAAA 302
Qy 198 -----198
Db 303 GGTTTACATATGTTGTCAGCATCTATTAAACAAATGATCGGATTCATTCACAACA 362
Qy 199 -----AsnHisValAlaGlnPheSerSerArgGlyProThrArg 211
Db 363 CAACCTCAATTTTATAATGAAATAATATATGGGATCATCTCATCAAAAGGTCACACAT 422
Qy 212 AspGlyArgIleLysProAspValMetAlaProGlyThrTyrIleLeuSerAlaArgSer 231
Db 423 GATGTTAGTGAACCTGATATAGTTGACCTGGTGAATATATTATACATCGGCAAGATCA 482
Qy 232 Ser-----LeuAlaProAspSerSerPheTrpAlaAsnHisAspSer 245
Db 483 AATGTCAGAAATTCACAGACCAATGTGGTGGTCTCTTTA-----CCAAATGCAAT 536
Qy 246 LysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIleValAlaGlyAsnValAla 265
Db 537 GGTCTAATGTCAATATCTGGTACATCAATGGCAACCATTTGGCAACAGCAGCAACA 596
Qy 266 GlnLeuArgGluHisPheVal 272
Db 597 ATCTTACACAAATATTAGTT 617

```

```

RESULT 10
BJ387574 601 bp mRNA linear EST 08-MAR-2002
LOCUS
DEFINITION
Dictyostelium discoideum cDNA library, SF Dictyostelium
ACCESSION
Dictyostelium discoideum cDNA clone ddc3a18 S', mRNA sequence.
VERSION
Dictyostelium discoideum
KEYWORDS
Dictyostelium discoideum
SOURCE
Dictyostelium discoideum
ORGANISM
Dictyostelium discoideum

```

Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.  
 1 (bases 1 to 601)  
 Urushihara H., Tanaka, Y., Kohara, Y. and Shin-i, T.  
 Full length cDNA of Dictyostelium discoideum at the slug stage  
 Unpublished (2002)  
 Contact: Tadasu Shin-i  
 Center For Genetic Resource Information  
 National Institute of Genetics  
 1111 Yata, Mishima, Shizuoka 411-8540, Japan  
 Tel: 81-559-81-6856  
 Fax: 81-559-81-6855  
 Email: tshin@genes.nig.ac.jp.  
 Location/Qualifiers  
 1. .601  
 /organism="Dictyostelium discoideum"  
 /mol\_type="mRNA"  
 /strain="AX4"  
 /db\_xref="taxon:44689"  
 /clone="dds3al8"  
 /sex="mat A"  
 /dev\_stage="slug stage"  
 /clone\_lib="Dictyostelium discoideum cDNA library, SF"

ORIGIN  
 Alignment Scores:  
 Pred. No.: 1.28e-08 Length: 601  
 Score: 190.50 Matches: 57  
 Percent Similarity: 44.86% Conservative: 26  
 Best Local Similarity: 30.81% Mismatches: 63  
 Query Match: 8.46% Indels: 39  
 DB: 12 Gaps: 6

US-09-985-689A-7 (1-433) x BJ387574 (1-601)

Qy 197 AsnLeuHisValAlaGlnPheSerArgGlyProThrArgAspGlyArgIleLys 216  
 |||||  
 Db 51 AATGAGAAATAATATTGTTCTCATCTCAAAAGGTCACACATGATGATGAGTAA 110  
 |||||  
 Qy 217 ProAspValMetAlaProGlyThrTyrlleLeuSerAlaArgSerIleuAla 234  
 |||||  
 Db 111 CCTGATTGTTGCCCTGTTGTAATATTACATCGCAGATCAATGTCCTACATA 170  
 |||||  
 Qy 235 -----ProAspSerPheThrAlaAsnHisAspSerLysTyrlaTyMet 250  
 |||||  
 Db 171 ACAGACCAATGTTGATGCTCTTTA---CCAAATACAAATCATATTGGCG---ATA 224  
 |||||  
 Qy 251 GlyClyThrSerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHis 270  
 |||||  
 Db 225 TCTGTTACATCAATGGCACTCATTTCGACGACGACCAACAATTTCTAGACAATAT 284  
 |||||  
 Qy 271 -----PheValIleAsnArgGlyValThrProLys 280  
 |||||  
 Db 285 TTAGTTGATGTTATATCCACTGGTTCAATTGTAGATCAATAATAATACCACT 344  
 |||||  
 Qy 281 ProSerLeuLysAlaAlaLeuIleAlaGlyAla----- 292  
 |||||  
 Db 345 GGATCAATATAAAGCATTAATGATTAATATGCTCAGTTATTAATGTCATCTCAA 404  
 |||||  
 Qy 293 -----AlaAspValGlyLeuGlyPheProAsnGlyAsn----- 303  
 |||||  
 Db 405 TTGATTACATCATCAAGTATTATCATATCCATCAACCAAGTTTGTGAATTTTCAGGT 464  
 |||||  
 Qy 304 -----GlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaPhe 319  
 |||||  
 Db 465 GCAAGTTTAGTTCAAGTGGGGTCTATTAGATGATGATGATGATGATGATGAT 524  
 |||||  
 Qy 320 ValAsnGlnThrSerProLeuSerThrSerClnLysAlaThrTyrlaThrAlaGln 339  
 |||||  
 Db 525 AATAAATAATAGTAATAATAATAAACAATCAATGATGATGATGATGATGAT 584  
 |||||  
 Qy 340 AlaGlyLysProLeu 344  
 |||||  
 Db 585 GGTGATTAGATTTA 599

RESULT 11  
 BQ770462  
 LOCUS  
 DEFINITION  
 ACCSSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM

BQ770462 665 bp mRNA linear EST 26-JUL-2002  
 UI-M-PI0-bvY-m-19-0-UI.r1 NIH BMAP\_F10 Mus musculus cDNA clone  
 IMAGE:5702970 5', mRNA sequence.  
 BQ770462 GI:21978936  
 EST.  
 Mus musculus (house mouse)  
 Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 665)  
 NIH-MGC http://mgi.nci.nih.gov/  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs@mail.nih.gov  
 Tissue Procurement: Dr. Jim Lin, University of Iowa  
 cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa  
 cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
 Clone Distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 This clone was contributed by the Brain Molecular Anatomy Project  
 (BMAP)

Seq primer: pYX-5.  
 Location/Qualifiers  
 1. .665  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:5702970"  
 /tissue\_type="whole brain"  
 /dev\_stage="embryo 12.5dpc"  
 /lab\_host="DH10B (T1 phage resistant)"  
 /clone\_lib="NIH BMAP\_F10"  
 /note="Organ: Brain; Vector: pYX-Asc; Site 1: EcoR I;  
 Site 2: Not I; The library was constructed according  
 Bonaldo, Lennon and Soares, Genome Research, 6:791-806,  
 1996. Denatured RNA was size fractionated on a 1% agarose  
 gel. First strand cDNA synthesis was primed with oligo-dT  
 primer containing a Not I site. Double strand cDNA was  
 size selected according to mRNA size fraction, ligated  
 with EcoR I adaptor, digested with NotI and then cloned  
 directionally into pYX-Asc vector. The library tag  
 sequence located between the Not I site and the polyA tail  
 is CAGCCACGAC. This library was created for the University  
 Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the  
 Developing Mouse Nervous System', supported by National  
 Institute of Mental Health (NIMH), Hemin Chin, Ph.D.,  
 program coordinator."

FEATURES  
 source

ORIGIN

Alignment Scores:  
 Pred. No.: 1.91e-08 Length: 665  
 Score: 189.50 Matches: 70  
 Percent Similarity: 43.51% Conservative: 34  
 Best Local Similarity: 29.29% Mismatches: 94  
 Query Match: 8.41% Indels: 41  
 DB: 13 Gaps: 9

US-09-985-689A-7 (1-433) x BQ770462 (1-665)

Qy 63 AsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnAlaThrAsn 82  
 |||||  
 Db 14 GATGATGGGTAGGGCCATGGCACATTCCTTCAGGTGTGATTCAGCATGAGGAGTGC 73  
 |||||  
 Qy 83 LysGlyMetAlaProGlnAlaAsnLeu---ValPheGlnSerIleMetAspSerGlyGly 101  
 |||||

```

Db      74 CAAGGATTGCTCCAGATGAGAGTGCACATCTTCCAGGCTTTTACCAACAAT----- 127
Qy      102 GlyLeuGlyGlyLeuProAlaAsnLeuGlnThrLeuPheSerGlnAlaTyrSerAlaGly 121
Db      128 -----CAGGTGCTTACACATCTTGGTTCTGGATGCTTCAATATATGCC 172
Qy      122 AlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyrThrAspSer 141
Db      173 ATCTTAAGAAGATGAGCTTCTCAACCTTACATCGGTGGCCGACCTTCATGATCAT 232
Qy      142 ArgAsnValAspAspTyr-----ValArgIleAsnAspMetThrIleLeuPheAlaAla 159
Db      233 CGTTTGTTCAGAGGTGGGAATACAGCTACAAATGAATGATGTTCTGCTATT 292
Qy      160 GlyAsnGluGlyProGlySerGlyThrIleSerAlaProGlyThrAlaIleAsnAlaIle 179
Db      293 GGCATATGAGGACCTCTCTATGCACTCTGAATAACCTGCTGATCAGATGATGATT 352
Qy      180 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 199
Db      353 GGAGTGGGTGGCATTGAC-----TTGAAGATAATACATC--- 385
Qy      200 HisValAlaAlaGlnPheSerSerArgGly-----ProThrArgAspGly 213
Db      386 -----GCTCGCTTTTCTCCAGGGAATGACTACCTGGGAATTACCAAGGAGGCTATGGT 439
Qy      214 ArgIleLeuProAspValMetAlaProGlyThrTyrIleLeuSerAlaArgSerSerLeu 233
Db      440 CGTGTGAAGCTGACATTGTC-----ACCTATGGTCTGGAGTGGGGGTTCCGGT 490
Qy      234 AlaProAspSerSerPheTrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThr 253
Db      491 GTGAAAGGGGGCTGTC-----CGTGCACTCTCAGGGACC 523
Qy      254 SerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPheValIys 273
Db      524 AGTGTGCTTCNCAGATGCTGCTGGGCGGCTGACCTTGTAGTAAAGCACAGTACAGAAG 583
Qy      274 AsnArgGlyValThrProLysProSerLeuLeuLeuIysAlaAlaLeuIleAlaGlyAla 292
Db      584 CGGAGCTGCTGTG-----AATCTGCCAGTGTGAAGCACGCTTTGATAGCTGACGCC 634

RESULT 12
CF727824
LOCUS
DEFINITION
UI-M-HB0-ckl-m-24-0-UI.r1 NIH_BMAP_H80 Mus musculus cDNA clone
IMAGE:30550487 5', mRNA sequence.
CF727824
EST.
CF727824.1 GI:37601992
Mus musculus (house mouse)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus;
1 (bases 1 to 675)
NIH-MGC http://mgi.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: Dr. James Lin University of Iowa
cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Distribution information can be found at
http://genome.uiowa.edu/distribution/mousef1.html
This genome was contributed by the Brain Molecular Anatomy Project
(BMAP).
Seq primer: pYX-5.
Location/Qualifiers
1..675
/organism="Mus musculus"
/mol_type="mRNA"
FEATURES
source

```

## ORIGIN

Alignment Scores:  
 Pred. No.: 3,93e-08 Length: 675  
 Score: 186.50 Matches: 69  
 Percent Similarity: 43.46% Conservative: 34  
 Best Local Similarity: 29.11% Mismatches: 93  
 Query Match: 8.28% Indels: 41  
 DB: 14 Gaps: 9

US-09-985-689A-7 (1-433) x CF727824 (1-675)

```

Qy      63 AsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnAlaThrAsn 82
Db      25 GAATGATGGGTAGGCGCATGGCCATCTGTCAGTGTGATGCCAGCATGAGGAGTGC 84
Qy      83 LysGlyMetAlaProGlnAlaAsnLeu---ValPheGlnSerIleMetAspSerGlyGly 101
Db      85 CAAGGATTTGCTCCAGATGCGAGCTGCACATCTTCAGGCTCTTACCAACAAT----- 138
Qy      102 GlyLeuGlyGlyLeuProAlaAsnLeuGlnThrLeuPheSerGlnAlaTyrSerAlaGly 121
Db      139 -----CAGGTGCTTACACATCTTGTGTTTCTGATGCTTCACTATGCC 183
Qy      122 AlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyrThrAspSer 141
Db      184 ATCTTAAGAAGATGAGCTTCTCAACCTTACATCGGTGGCCGACCTTCATGATCAT 243
Qy      142 ArgAsnValAspAspTyr-----ValArgIleAsnAspMetThrIleLeuPheAlaAla 159
Db      244 CGTTTGTTCAGAGGTGGGAATTAACAGCTAACCAATGATATATGTTTCTGCTATT 303
Qy      160 GlyAsnGluGlyProGlySerGlyThrIleSerAlaProGlyThrAlaIleAsnAlaIle 179
Db      304 GGCATATGAGGACCTCTCTATGCACTCTGAATAACCTGCTGATCAGATGATGATT 363
Qy      180 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 199
Db      364 GGAGTGGGTGGCATTGAC-----TTGAAGATAATACATC--- 396
Qy      200 HisValAlaGlnPheSerSerArgGly-----ProThrArgAspGly 213
Db      397 -----GCTCGCTTTTCTCCAGGGAATGACTACCTGGGAATTACCAAGGAGGCTATGGT 450
Qy      214 ArgIleLysProAspValMetAlaProGlyThrTyrIleLeuSerAlaArgSerSerLeu 233
Db      451 CGTGTGAAGCTGACATTGTC-----ACCTATGGTCTGGAGTGGGGGTTCCGGT 501
Qy      234 AlaProAspSerSerPheTrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThr 253
Db      502 GTGAAAGGGGGCTGTC-----CGTGCACTCTCAGGGACC 534

```

/strain="C57BL/6"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:30550487"  
 /tissue\_type="whole eye"  
 /dev\_stage="embryo 12.5,13.5,14.5 dpc"  
 /lab\_host="DH10B (T1 phage resistant)"  
 /clone\_lib="NIH\_BMAP\_H80"  
 /note="Organ: Eye; Vector: pYX-Asc; Site: 1: EcoR I;  
 Site 2: Not I; The library was constructed according  
 Bonaldo, Lennon and Soares, Genome Research, 6:791-806,  
 1996. Denatured RNA was size fractionated on a 1% agarose  
 gel. First strand cDNA synthesis was primed with oligo-dT  
 primer containing a Not I site. Double strand cDNA was  
 size selected according to mRNA size fraction, ligated  
 with EcoR I adaptor, digested with NotI and then cloned  
 directionally into pYX-Asc vector. The library tag  
 sequence located between the Not I site and the polyA tail  
 is TATATGAGT. This library was created for the University  
 Iowa Brain Anatomy Project (BMAP). Gene Discovery in the  
 Developing Mouse Nervous System, supported by National  
 Institute of Mental Health (NIMH)."



Thomas L. Casavant.  
Web site: <http://genome.uiowa.edu>; tom-casavant@uiowa.edu  
Contact: bentto-soaresuiowa.edu; Bair, T., Bair, J., Crouch, K., Davis, A.,  
Bonaldi, M.F., Akabogu, I., Bair, T., Bair, J., Crouch, K., Davis, A.,  
Fisher, K., Keppel, C., Kucaba, T., Lebeck, M., Melo, A., Schaefer, K.,  
Scheetz, T., Smith, C., Snir, E., Tack, D., Trout, K., Walters, J.,  
Casavant, T., Soares, M.B.

Clone distribution: MGC clone distribution information can be found  
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Series: Plate: Row: Column: 0  
This clone was selected for full length sequencing because it  
passed the following selection criteria: matched mRNA gi: 6678418  
This clone has the following problem: frame shifted.

Location/Qualifiers  
1. 4662  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6"  
/db\_xref="taxon:10090"  
/clone="IMAGE:684459"  
/tissue\_type="Brain, mouse, 13.5,14.5,16.5,17.5 dpc"  
/clone\_lib="NIH\_BMAP\_FYO"  
/lab\_host="DH10B"  
/note="vector: pyx-ASC"

# ORIGIN

Alignment Scores:  
Pred. No.: 2,99e-06 Length: 4662  
Score: 182.00 Matches: 102  
Percent Similarity: 36.54% Conservative: 69  
Best Local Similarity: 21.79% Mismatches: 159  
Query Match: 8.08% Indels: 138  
DB: 11 Gaps: 22

US-09-985-689A-7 (1-433) x BC060627 (1-4662)

QY 10 AlaaspValAlaGlnAsnAspPheGlyLeuTyrglyGlnGlyGlnIleValAlaValAla 29  
DB 803 GCTGAGATGCTGAATTAATCTGTGAACATTTATGACGATGGAAACCTGCTCTCCATTTGTG 862  
QY 30 AspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisGluAlaPheArgGlyLys 49  
DB 863 ACCACCGA----- 871  
QY 50 IleThrAlaLeuTyAlaLeuGlyArgThrAsnAsnAlaAsnAspProAsnGlyHisGly 69  
DB 872 -----GCAGCTCATGGA 883  
QY 70 ThrHisValAlaGlySerValLeuGlyAsn-----AlaThrAsnLysGlyMet 85  
DB 884 ACCCATGTAGCAAGTATAGCCCGCAGGCGATTTTCCAGAGAGCCTGGAACGGAATGGAGTT 943  
QY 86 AlaProGlnAlaAsnLeuValPhe-----GlnSerIleMetAsp 98  
DB 944 GCTCTGGTGTCAAAATTCATCCATTAAGATTGGTGATACACGGCTAAGCACTATGGAA 1003  
QY 99 SerGlyGlyLeuGlyLeuProAlaAsnLeuGlnThrLeuPheSerGlnAla--- 117  
DB 1004 ACAGGCACAGGCTC-----ATCAGAGCTATGATAGAGTTATTAATCATAGTGTAT 1057  
QY 118 -----TySerAlaGlyAlaArgIleHisThrAsnSerTTPGlyAlaProValAsn 134  
DB 1058 CTTGTCAACTACATGTTATGGAGAGCAACTCAT-----TGG-----CCAAATTTCT 1102  
QY 135 GlyAlaTyThrThrAspSerArgAsnValAspAspTyValArgLysAsnAspMetThr 154  
DB 1103 GGAGAGAATT-----TGTGAAGTAATTAATGAAGCAGTATGGAACATTAATCAAT 1153  
QY 155 IleLeuPheAlaAlaGlyAsnGluGlyProGlySerGlyThrIleSerAlaPro---Gly 173  
DB 1154 TATGTTTCAAGTCTGGAAATATATGTCATGCTTTCTACAGTGGTGTGTCACAGGAGA 1213

QY 174 ThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsnLeuArgProSerPheGlySer 193  
DB 1214 ACTACATCCAGTGTGTAGGTGTGGAGCT----- 1243  
QY 194 TyrAlaAspIleAsnHisValAlaGlnPhe----- 204  
DB 1244 TATGTTTCCCTGATATGATGTTGCAGAGTATTCATCTGAGAGAGAACTGCCTGCAAT 1303  
QY 205 -----SerSerArgGlyProThrArgAspGlyArgIleLysProAspValMet 220  
DB 1304 CAATATACATGGTCTTTAGAGGCCCAAGTGTGATGGAGCCCTCGGTGTGAGCATCAGT 1363  
QY 221 AlaProGlyThrTyIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTip 240  
DB 1364 GCACAGGAGGTGTATTTGCTTCTGCTGCTAAC-----TGG 1399  
QY 241 AlaAsnHisAspSerLysTyAlaTyMetGlyGlyThrSerMetAlaThrProIleVal 260  
DB 1400 ACATTTAGGGGGAGCTCAG-----CTAATGAATGGGACATCAATGTCTCCCCCAATGCC 1453  
QY 261 AlaGlyAsnValAla-----GlnLeuArgGluHisPheValLysAsnArgGly 276  
DB 1454 TGTGTGGCATTTGCCCTGTGTTTTCAGGGCTGAAGCAATAATATGTGACTACTACTGTA 1513  
QY 277 ValThrProLysProSerLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspValGly 296  
DB 1514 CACTCAGTCAGAGAGCTCTAGAAATATCTGCAATAAAGCTGACAAATATAGAGTATT 1573  
QY 297 LeuGlyPheProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsn 316  
DB 1574 GCC-----CAAGGACATGGAATTTATTCAGGTGTGACAAAGCTTATGAC 1615  
QY 317 ValAlaPheValAsnGluThrSerProLeuSerThrSerGlnLysAlaThrTySerPhe 336  
DB 1616 TACCTCATTCAAAAT-----ACATCATTTGCTAACAGATAGGTTT 1657  
QY 337 ThrAlaGlnAlaGlyLys-----ProLeuLysIleSerLeu 348  
DB 1658 ACAGTTACTGTTGGAATAAACCGTGTATCTACCTCCGAGATCTCTGTCAGGTGGCTGCT 1717  
QY 349 ValTrpSerAsp-----AlaProGlySerThrThrAlaSer 360  
DB 1718 CCT---TCAGATCATGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1774  
QY 361 LeuThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrLysTyVal 380  
DB 1775 GAAAAATATCTTTTCAGCTTTCATTTAGCTTTAACTTCA----- 1813  
QY 381 GlyAspAspPheThrAlaProTyAspAsnAsnTrpAspGlyArgAsnAsnValGluAsn 400  
DB 1814 ---AATTCATCTTGGGTTTCAGTGTCCCGCCATTTTGAACCTCATGAACTGTCGGCAC 1870  
QY 401 ValPheIleAsnAlaPro-----GlnSerGlyThrTyThrValGluValGln 416  
DB 1871 ATAAACATACGTTGGAGCCCGCCAGGGGCTTAAGAGAGGGTTTACATTATACAGAGTATGT 1930  
QY 417 AlaTyAsnValProValSerPro 424  
DB 1931 GGCTATGATATA---GCATCCCCC 1951  
RESULT 15  
BC011275 3091 bp mRNA linear HTC 19-NOV-2003  
LOCUS Mus musculus tripeptidyl peptidase II, mRNA (cDNA clone  
IMAGE:3674891), containing frame-shift errors.  
DEFINITION  
ACCESSION BC011275  
VERSION BC011275.1 GI:15030057  
KEYWORDS HTC  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 3091)

## AUTHORS

Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Altshul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hejeh, P., Diachenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Cabavant, T.L., Schetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S., Carrinci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullaly, S.J., Bosak, S.A., McSwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahy, J., Helton, E., Kettman, M., Madan, A., Rodrigues, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakeley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smal, D.E., Scherch, A., Schein, J.E., Jones, S.J., and Marra, M.A.

Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences  
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

2 (bases 1 to 3091)  
12477932  
PubMed  
Strausberg R.  
Direct Submission  
Submitted (25-JUL-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA  
NTH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: MGC help desk  
Email: [gcaps-rc@mail.nih.gov](mailto:gcaps-rc@mail.nih.gov)  
Tissue Procurement: Gilbert Smith, Ph.D.  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center  
Center code: BCM-HGSC  
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>  
Contact: [amg@bcm.tmc.edu](mailto:amg@bcm.tmc.edu)  
Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Loulseged, H., Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati, A.N., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Series: IRAC Plate: 23 Row: 1 Column: 4  
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 6678418  
This clone has the following problem: frame shifted.

## FEATURES

## source

1..3091  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="Czech 11"  
/db\_xref="taxon:10090"  
/clone="IMAGE:3674891"  
/tissue types="Mammary tumor metastasized to lung. Tumor arose spontaneously from a senescent normal mammary (clonal) outgrowth infected with the virus MMV."  
/clone\_lib="NCI-CGAP\_Lu29"  
/lab\_host="DH10B"  
/note="vector: pCMV-SPORT6"

## ORIGIN

Alignment Scores:  
Pred. No.: 1.87e-06 Length: 3091  
Score: 181.00 Matches: 84  
Percent Similarity: 37.26% Conservative: 52  
Best Local Similarity: 23.01% Mismatches: 121  
Query Match: 8.04% Indels: 108  
DB: 11 Gaps: 16

US-09-985-689A-7 (1-433) x BC011275 (1-3091)

QY 10 AlaAspValAlaGlnAsnAsnPheGlyLeuTyrglyGlnGlyGlnIleValAlaValAla 29  
Db 771 GCTGAGATGCTGAATTAATCTCTGTGAACATTATGACGATGGGAACCTGCTCCATGTG 830  
QY 30 AspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisGluAlaPheArgGlyLys 49  
Db 931 ACCAGCGGA----- 839  
QY 50 IleThrAlaLeuTyralaLeuGlyArgThrAsnAsnAlaAsnAspProAsnGlyHisGly 69  
Db 840 -----GGAGCTCATGGA 851  
QY 70 ThrHisValAlaGlySerValLeuGlyAsn-----AlaThrAsnLysGlyMet 85  
Db 852 ACCCATGTAGCAAGTATAGCGGAGGCATTTTCCAGAGAGGCTGAAACGGAATGGAGTT 911  
QY 86 AlaProGlnAlaAsnLeuValPhe-----GlnSerIleMetAsp 98  
Db 912 GCTCCTGGTGTCAAATTCATCCATTAAGATTGGTGATACACGGCTTAACACTATGGAA 971  
QY 99 SerGlyGlyLeuGlyGlyLeuProAlaAsnLeuGlnThrLeuPheSerGlnAla--- 117  
Db 972 ACAGGCACAGGCCCTC-----ATCAGAGCTATGATAGAAGTTATAAATCATAGTGTGAT 1025  
QY 118 -----TyrSerAlaGlyAlaArgIleHisThrAsnSerTTPGlyAlaProValAsn 134  
Db 1026 CTTGTCAACTACAGTTATGGAGAACCACTCAT-----TGG-----CCAAATCT 1070  
QY 135 GlyAlaTyThrThrAspSerArgAsnValAspAspTyrValArgLysAsnAspMetThr 154  
Db 1071 GGGAGAATT-----TGTGAAGTAATTAATGAAGCAGTATGGAACATAATAACAATT 1121  
QY 155 IleLeuPheAlaAlaGlyAsnGluGlyProGlySerGlyThrIleSerAlaPro---Gly 173  
Db 1122 TATGTTTCTAGTGTCTTCTAGAGGCCCAAGTCTGATGGAGCCCTCGGTGTCAGCATCAGT 1181  
QY 174 ThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsnLeuArgProSerPheGlySer 193  
Db 1182 ACTACATCCAGTGTAGTAGTGTGGAGCT----- 1211  
QY 194 TyrAlaAspAsnIleAsnHisValAlaGlnPhe----- 204  
Db 1212 TATGTTTCCCTGATATGATGTTTGCAGAGTATTCACCTAGAGAGAGAACTGCCTGCAAT 1271  
QY 205 -----SerSerArgGlyProThrArgAspGlyArgGlyLysProAspValMet 220  
Db 1272 CAATATACATGTTCTTCTAGAGGCCCAAGTCTGATGGAGCCCTCGGTGTCAGCATCAGT 1331  
QY 221 AlaProGlyThrTyrlleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTTP 240  
Db 1332 GCACCCAGGAGGTGCTTATTGCTTCTGTGCCTAAC-----TGG 1367  
QY 241 AlaAsnHisAspSerLysTyralaTyMetGlyGlyThrSerMetAlaThrProIleVal 260  
Db 1368 ACATTGAGGGGGAGCTCAG-----CTAATGAATGGACATCAATGCTCTCCCCAATGCC 1421  
QY 261 AlaGlyAsnValAla-----GlnLeuArgGluHisPheValLysAsnArgGly 276  
Db 1422 TGTGTTGGCATTCCTGCTTCTTTCAGGCTGAAAGCAATAATATGTTGACTACTGTA 1481  
QY 277 ValThrProLysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspValGly 296  
Db 1482 CACTCAGTCAGAGAGAGCTCTAGAAAATACTGCAATAAAAGTGCACATATAGAAGTATT 1541  
QY 297 LeuGlyPheProAsnGlyAsnGlnGlyTTPGlyArgValThrLeuAspLysSerLeuAsn 316  
Db 1542 GCC-----CAAGGACATGAATATTATTCAGTTGACAAAGCTTATGAC 1583  
QY 317 ValAlaPheValAsnGluThrSerProLeuSerThrSerGlnLysAlaThrTyrSerPhe 336



Db 1584 TACTCATTCARAAAT-----ACATCATTTGCTAACAGATTAGGTTTT 1625

Qy 337 ThrAlaGlnAlaGly 341

Db 1626 ACAGTTACTGTGGA 1640

Search completed: April 4, 2004, 11:55:24  
Job time : 2286.58 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: April 3, 2004, 16:37:54 ; Search time 335.908 Seconds  
(without alignments)  
5476.111 Million cell updates/sec

Title: US-09-985-689A-7  
Perfect score: 2252  
Sequence: 1 NDVARGIVKADVAQNNGFLY.....EVOXNVVPSPOTFSLAIVH 433

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 3373863 seqs, 212409041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPRO.spool/US0985689/runat\_31032004\_161806\_4145/app\_query.fasta\_1.3498  
-DB=N Geneseg 29Jan04 -Qfmt=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.cdi  
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFWT=200 -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US0985689 -CGN\_1\_1238 @runat\_31032004\_161806\_4145 -NCPU=6 -ICPU=3  
-NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N Geneseg 29Jan04:\*  
1: Geneseg1980s:\*  
2: Geneseg1990s:\*  
3: Geneseg2000s:\*  
4: Geneseg2001as:\*  
5: Geneseg2001bs:\*  
6: Geneseg2002as:\*  
7: Geneseg2003as:\*  
8: Geneseg2003bs:\*  
9: Geneseg2003cs:\*  
10: Geneseg2004s:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2252	100.0	3003	AAV82382	AAV82382 Bacillus
2	2125.5	94.4	1923	AAV82382	AAV82382 Bacillus
3	2120.5	94.2	1923	AAV82382	AAV82382 Bacillus
4	2110.5	93.7	1920	AAV82382	AAV82382 Bacillus
5	2032	90.2	1299	AAQ27516	AAQ27516 Alkali-pr
6	432.5	19.2	1977	AAV85667	AAV85667 Thermococ
7	432.5	19.2	1977	AAV85667	AAV85667 Thermococ
8	396	17.6	1236	AAV85920	AAV85920 Hyperther

9	396	17.6	1566	2	AAV85668	Aat85668 Pyrococcus
10	396	17.6	1962	2	AAV85668	Aat85668 Pyrococcus
11	396	17.6	1962	2	AAV85668	Aat85668 Pyrococcus
12	390	17.3	1977	2	AAV85669	Aat85669 Hyperther
13	369.5	16.4	2121	4	ABL54900	ABL54900 T. yonsei
14	327.5	14.5	2539	2	AAV81454	Aat61454 Streptomy
15	327.5	14.5	2809	2	AAV81455	Aat61455 DhpA-mel
16	325	14.4	1329	6	ABK74643	ABK74643 Bacillus
17	323	14.3	135638	7	ABX34289	ABX34289 S. atrool
18	311.5	13.8	2835	2	AAV81411	Aat08141 Hyperther
19	311.5	13.8	4765	2	AAV81412	Aat08142 Hyperther
20	311.5	13.8	4765	2	AAV85670	Aat85670 Pyrococcus
21	311.5	13.8	4765	2	AAV85671	Aat85671 WO9856926
22	309	13.7	2532	2	AAQ23134	Aaq23134 Encodes R
23	297	13.2	1560	7	ABZ37569	ABZ37569 Streptomy
24	297	13.2	1859	2	AAV85677	Aat85677 Thermococ
25	297	13.2	59816	7	ABZ37516	ABZ37516 Streptomy
26	297	13.2	59816	7	ABZ37516	ABZ37516 Streptomy
27	289	12.8	3413	2	AAV72330	AAV72330 F. balust
28	278.5	12.4	3743	9	ADD24901	Add24901 DNA encod
29	278.5	12.4	546	4	ABL53453	ABL53453 T. yonsei
30	276.5	12.3	3788	9	ADD24905	Add24905 DNA encod
31	276	12.3	898	2	AAV8131	Aat08131 Hyperther
32	275.5	12.2	4716	9	ABQ80437	ABQ80437 CSPa codi
33	275.5	12.2	4740	7	ABZ58957	ABZ58957 Group B S
34	271.5	12.1	1306	6	ABL55784	ABL55784 Bacillus
35	271.5	12.1	1330	6	ABL55784	ABL55784 Bacillus
36	270	12.0	564	2	AAV8134	AAV8134 DNA seque
37	270	12.0	564	2	AAV85676	Aat85676 Thermococ
38	268.5	11.9	4650	6	ABN71526	ABN71526 Streptoco
39	268.5	11.9	4710	6	ABN71162	ABN71162 Streptoco
40	268.5	11.9	4734	6	ABN70525	ABN70525 Streptoco
41	268.5	11.9	4770	6	ABN69191	ABN69191 Streptoco
42	268.5	11.9	110000	6	ABN71527_19	Continuation (20 o
43	258	11.5	2166	6	ABD34433	ABD34433 C-termina
44	258	11.5	2166	6	ABL40498	ABL40498 DNA const
45	256.5	11.4	807	2	AAQ98740	Aaq98740 DNA encod

ALIGNMENTS

RESULT 1  
AAV82382  
ID AAV82382 standard; DNA; 3003 BP.

XX AC AAV82382;  
XX 12-APR-1999 (first entry)  
XX DE Bacillus JF170 protease gene.

XX KW Protease; detergent; surfactant; leather processing; debittering;  
XX KW flavour; ss.  
XX OS Bacillus sp.

XX FH Key Location/Qualifiers  
XX FT CDS 846..2771  
XX FT sig\_peptide /\*tag= a  
XX FT sig\_peptide /\*tag= b  
XX FT mat\_peptide 1470..2768  
XX FT /\*tag= c

XX PN WO9856927-A2.

XX PD 17-DEC-1998.

XX PF 09-JUN-1998; 98WO-US012005.

XX PR 12-JUN-1997; 97US-00873479.

XX PA (NOVO ) NOVO NORDISK BIOTECH INC.

1020

XX Sloma A, Christianson L;  
XX WFI; 1999-080908/07.  
DR P-PSDB; AAW89547.  
XX  
XX Novel protease from *Bacillus subtilis* LC20 - useful in laundry and  
XX dishwashing detergents and for leather processing.  
XX  
XX Claim 11; Page 52-53; 77pp; English.  
XX  
XX This nucleotide sequence encodes a novel protease (see AAW82382) of  
CC *Bacillus* sp. JPI170 (NCIB 12513). The sequence in plasmid p170BAN is  
CC contained in *Bacillus subtilis* LC20 NRRL B-21680. The protease gene was  
CC isolated from chromosomal DNA of JPI170 following preparation of probes  
CC based on protease N-terminal and internal peptides (see AAW89549-50),  
CC screening of chromosomal libraries, isolation of the 3' end of the gene  
CC by inverse PCR (see AAW82410-11), reconstruction of 5' and 3' ends and  
CC PCR amplification (see AAW82412-16). Claimed recombinant host cells can  
CC be used in a method for producing the protease. The protease is used in  
CC laundry and dishwashing detergents, for institutional and industrial  
CC cleaning, and for leather processing, as well as for debittering and  
CC enhancing the degree of hydrolysis of protein hydrolysates, for flavour  
CC development through hydrolysis of proteins, degradation of undesired  
CC peptides and in enzymatic synthesis of peptides. It has enhanced  
CC stability towards oxidation under alkaline conditions, e.g. towards  
CC bleaching agents of the peroxy type. The invention also provides mutant  
CC cells in which the protease activity is diminished. Such cells can be  
CC used for the production of heterologous recombinant proteins  
XX  
SQ Sequence 3003 BP; 982 A; 504 C; 645 G; 872 T; 0 U; 0 Other;

Alignment Scores:  
Pred. No.: 4.81e-170 Length: 3003  
Score: 2252.00 Matches: 433  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 2 Gaps: 0

US-09-985-689a-7 (1-433) x AAW82382 (1-3003)

QY 1 AsnAspValAlaArgGlyIleValLysAlaAspValAlaGlnAsnAsnPheGlyLeuTyr 20  
DB 1470 AATGACGTGGCCGTTGGCAATGTGAAGCAGACGTCGCACAAAATAACTTTGGCTTATAT 1529

QY 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40  
DB 1530 GGACAGGACAGATTGTAGCAGTTCTGATCTGGCTTGATACAGGAGAAATGACAGT 1589

QY 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60  
DB 1590 TCGATGCATGAAGCATTCGCGGTAAAGATTACCGCACATATATGCACCTGGCGAGAACGAAT 1649

QY 61 AsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnAla 80  
DB 1650 AACGCCAATGATCCAAATGGACATGGACCCCATGTGCTGGATCTGTGTAGGAATGCT 1709

QY 81 ThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSerGly 100  
DB 1710 ACAATTAAGGATGGCACCAGCCCAATCTAGCTTTCAATCTATTATGATAGTGGT 1769

QY 101 GlyGlyLeuGlyGlyLeuProAlaAsnLeuGlnThrLeuPheSerGlnAlaTyrSerAla 120  
DB 1770 GGAGGCGCTGGGAGGACTACTCTGTAATCTACAAACATTATTTCAGTCAAGCATATAGTCT 1829

QY 121 GlyAlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyValaTyrThrThrAsp 140  
DB 1830 CGAGCAGAAATTCATACGATTCATGGGGGCTCCAGTAAGCGTGCCTATACGACAGAC 1889

QY 141 SerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAlaGly 160  
DB 1890 TCTCGAAATGTTGATGATTATGTGAGAAAAATGATATGACGATTCCTTTTTCGGCCCGGA 1949

QY 161 AsnGluGlyProGlySerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThr 180  
DB 1950 AATGAGGACCCAGGTAGCGGTACAAATCAGTGCACGAGAAACAGAAAAATGCGATTACA 2009

QY 181 ValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsnHis 200  
DB 2010 GTTGGGGACACCGAAACCTACGTCGAAGCTTCGATCTTATGGGATATATTAAACCAT 2069

QY 201 ValAlaGlnPheSerSerArgGlyProThrArgAspGlyArgIleLysProAspValMet 220  
DB 2070 GTTGCTCAATCTCTTCACGAGGTCCTACTAGATGACGATATTAAAGCCGAGCGTCATG 2129

QY 221 AlaProGlyThrTyrIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrp 240  
DB 2130 GCACGAGTACGTATATTTCTCTCTGATAGATCATATTAGTCCAGATTCCTCATTTCTGG 2189

QY 241 AlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIleVal 260  
DB 2190 GCAAAACCATGATAGTAAATATGCTACATGCGTGGTACTTCTATGGCTACTCCAAATGTA 2249

QY 261 AlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyValThrProLys 280  
DB 2250 GCAGTAATGTTGCAAAATTAAGGGAGCATTTTGTGAAAAATAGAGGGGTAACTCCCTAAG 2309

QY 281 ProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAspValGlyLeuGlyPhePro 300  
DB 2310 CCTTCCCTTTTAAAGCTGCTTTAATTCAGGTGCTGGGATGTTGGACTTGGCTTTCCA 2369

QY 301 AsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaPheVal 320  
DB 2370 AATGGTAACCAAGGATGGGAAGTAACGTTAGATAAATCCCTAAATGTGCATTTGTG 2429

QY 321 AsnGluThrSerProLeuSerThrSerGlnLysAlaThrTyrSerPheThrAlaGlnAla 340  
DB 2430 AATGAACAGAGCCCTTTATCAACAGTCAAAAAGCAACATATTCGTTTACGGCTCAAGCT 2489

QY 341 GlyLysProLeuLysIleSerLeuValTrpSerAspAlaProGlySerThrThrAlaSer 360  
DB 2490 GGTAAACCCCTTAAAAATATCACTTGTGTCAGATGCACCGAGTAGCACCGGCATCA 2549

QY 361 LeuThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrLysTyrVal 380  
DB 2550 CTAACCTTTAGTGAATGATTTAGACTTAGTATCATCTGCACCAATGGAATAAATACGTC 2609

QY 381 GlyAsnAspPheThrAlaProTyrAspAsnAsnTrpAspGlyArgAsnAsnValGluAsn 400  
DB 2610 GGAATATGACTTTTACAGCACCGTATGATACAAATGGGATGGCAGAAACACACGCTGAAAT 2669

QY 401 ValPheIleAsnAlaProGlnSerGlyThrTyrThrValGluValGlnAlaTyrAsnVal 420  
DB 2670 GTGTTTATCAATGCTCTCTCAAGCGGAGAGTATACAGTCGAAGTGCAGGCTTACAATGTA 2729

QY 421 ProValSerProGlnThrPheSerLeuAlaIleValHis 433  
DB 2730 CCAGTAAGTCCGCAAAACCTTTCTTTAGCGATTGTACAT 2768

RESULT 2  
AA37278 standard; DNA; 1923 BP.  
ID AAX37278  
XX AC AAX37278;  
XX  
XX 20-MAR-2003 (revised)  
DT 21-JUL-1999 (first entry)  
XX  
XX *Bacillus* alkaline protease encoding DNA.  
XX Alkaline protease; *Bacillus*; casein digestion; oleic acid; enzyme;  
KW washing composition; oxidising agent; ss.  
XX *Bacillus* sp.  
OS  
XX

PN WO9918218-A1.  
 XX 15-APR-1999.  
 XX 07-OCT-1998; 98WO-JP004528.  
 XX 07-OCT-1997; 97JP-00274570.  
 XX (KAOS ) KAO CORP.  
 XX Takaiwa M, Okuda M, Saeki K, Kubota H, Hitomi J, Kageyama Y;  
 PI Shikata S, Nomura M;  
 XX WPI; 1999-287736/27.  
 DR P-PSDB; AAY17088, AAY17090.  
 XX Alkali protease from Bacillus used in washing powders.  
 XX Disclosure; Page 58-63; 71pp; Japanese.  
 CC The invention relates to alkaline proteases produced by strains of  
 CC Bacillus. The proteases ability to digest casein is not inhibited by  
 CC oleic acid and they have a high stability to oxidising agents. The  
 CC alkaline protease of the invention has the following properties: (a) it  
 CC is active over the pH range 4-13 and has at least 80% of its optimum  
 CC activity over the range pH 6-12; (b) after 30 minutes at 40 deg. C it is  
 CC stable over the pH range 6-11; (c) its isoelectric point is 8.9-9.1; (d)  
 CC its ability to digest casein is not inhibited by oleic acid; (e) it has  
 CC molecular weight about 43,000 by SDS-PAGE. The alkaline proteases can be  
 CC used as enzymes in washing compositions for use in automatic dishwashers  
 CC and for washing clothes. The stability to oxidising agents allows the  
 CC enzyme to be an effective component of washing compositions including  
 CC bleaches. The present sequence represents an alkaline protease encoding  
 CC DNA. (Updated on 20-MAR-2003 to correct DR field.)  
 XX  
 SQ Sequence 1923 BP; 576 A; 418 C; 475 G; 454 T; 0 U; 0 Other;

Alignment Scores:  
 Pred. No.: 3,47e-160 Length: 1923  
 Score: 2125.50 Matches: 406  
 Percent Similarity: 97.93% Conservative: 19  
 Best Local Similarity: 93.55% Mismatches: 8  
 Query Match: 94.38% Indels: 1  
 DB: 2 Gaps: 1

US-09-985-689A-7 (1-433) x AAX37278 (1-1923)

Qy 1 AsnAspValAlaArgGlyIleValLysAlaAspValAlaGlnAsnAsnPhedGlyLeuTyr 20  
 Db 619 AATGATCTTCGGGTGGAATGTGCAAGCGGATGTGGCTCAGAGCAGCTACGGGTGTAT 678  
 Qy 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40  
 Db 679 GGACAGGACAGATCGTAGCGGTTCGCGATACAGGGCTTGATCAGGTGCAATGACAGT 738  
 Qy 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60  
 Db 739 TCGATGATGAAGCCTTCGCGGGGAAATTTACTGCATTATATCATTTGGACCGGCAAT 798  
 Qy 61 AsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn-- 79  
 Db 799 AATGCCAATGATACGAATGTGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTC 858  
 Qy 80 AlaThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 99  
 Db 859 TCCACTAATAAGGAATGGGCGCTCAGCGGAATCTAGTCTTCCAAATCTATCATGATGATG 918  
 Qy 100 GlyGlyGlyLeuGlyLeuProAlaAsnLeuGlnThrLeuPheSerGlnAlaTyrSer 119  
 Db 919 GGTGGGGACTTGGAGGACTACCTTGAATCTGAAATCTGAAACCTTATTCAGCAAGCATACAGT 978  
 Qy 120 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyrThrThr 139

Db 979 GCTGGTCCAGAAATTCATACAAACTCTCTGGGGAGCAGCAGTGAATGGGGCTTACACAACA 1038  
 Qy 140 AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 159  
 Db 1039 GATTCAGAAATGTGGATGACTATGTGCGCAAAATGATATGACGATCCTTTTCGTGCC 1098  
 Qy 160 GlyAsnGluGlyProGlySerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 179  
 Db 1099 GGGAAATGAAGACCGAAGCGGGAACCATCAGTGCCACCGCAGCAGTAAATATGCAATA 1158  
 Qy 180 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 199  
 Db 1159 ACAGTCGGAGCTACGGAAACCTCCGCGCAAGCTTTGGGTCTTATGCGGCAATATCAAC 1218  
 Qy 200 HisValAlaGlnPheSerSerArgGlyProThrArgAspGlyArgIleLysProAspVal 219  
 Db 1219 CATGTGCACAGTTCTCTTCAGTGGACCGCAACGAATGGACGGATCAACACCGGATGTC 1278  
 Qy 220 MetAlaProGlyThrTyrIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 239  
 Db 1279 ATGGCACCGGAACGTTCTATCTATCATCAGCAAGATCTTCTCTTGCACCGGATTCCTCTC 1338  
 Qy 240 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProIle 259  
 Db 1339 TGGCGCAACCATGACAGTAAATATGATATGATGCGTGAACGCTCCATGGCTTACACCGATC 1398  
 Qy 260 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyValThrPro 279  
 Db 1399 GTTGTGGAACGTTGGACAGCTTCGTGAGCATTTTGTGAAACACAGGAGCATCACACCA 1458  
 Qy 280 LysProSerLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspValGlyLeuGlyPhe 299  
 Db 1459 AAGCCTTCTCTATTAAAGCGGCACTGATTCGCGGTGCGAGCTGACATCGGCTTGGGTAC 1518  
 Qy 300 ProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaPhe 319  
 Db 1519 CCGAACGGTAACCAAGGATGGGACGAGTGATGATGATGATGATGATGATGATGATGATGAT 1578  
 Qy 320 ValAsnGluThrSerProLeuSerThrSerGlnLysAlaThrTyrSerPheThrAlaGln 339  
 Db 1579 GTGAACGAGTCCAGTTCTCTATCCACCGCCCAAAAGCGAGCTACTCGTTTACTGTACT 1638  
 Qy 340 AlaGlyLysProLeuLysIleSerLeuValTrpSerAspAlaProGlySerThrAla 359  
 Db 1639 GCGGCAAGCCTTTGAAATCTCCCTGGTATGGTCTGATGCCCTGCGAGCAGCACTGCT 1698  
 Qy 360 SerLeuThrLeuValAsnAspLeuValIleThrAlaProAsnGlyThrLysTyr 379  
 Db 1699 TCCGTAAACGCTTGTCATGATCTGGACCTTCTCATACCGCTCCAAATGGCACACAGTAT 1758  
 Qy 380 ValGlyAsnAspPheThrAlaProTyrAspAsnAsnTrpAspGlyArgAsnAsnValGlu 399  
 Db 1759 GTAGGAAATGACTTTTACTTCGCCATACATTAATCTGGGATGGCGCAATAACGTAGAA 1818  
 Qy 400 AsnValPheIleAsnAlaProGlnSerGlyThrTyrValGluValGlnAlaTyrAsn 419  
 Db 1819 AATGATTTTATTAATGACCAACAGCGGACGATATACATTAATGAGGTACAGGCTTATAC 1878  
 Qy 420 ValProValSerProGlnThrPheSerLeuAlaIleValHis 433  
 Db 1879 GTACCGGTGGACCAACAGACCTTCTCGTTGGCAATTGTGAAT 1920

RESULT 3  
 AAX37279  
 ID AAX37279 standard; DNA; 1923 BP.  
 XX  
 AC AAX37279;  
 XX  
 XX 20-MAR-2003 (revised)  
 DT 21-JUL-1999 (first entry)  
 XX  
 DE Bacillus alkaline protease encoding DNA.  
 XX

KW Alkaline protease; Bacillus; casein digestion; oleic acid; enzyme;  
 XW washing composition; oxidising agent; ss.  
 XX Bacillus sp.  
 OS WO918218-A1.  
 XX 15-APR-1999.  
 PD 07-OCT-1998; 98WO-JP004528.  
 XX 07-OCT-1997; 97JP-00274570.  
 XX (KAO ) KAO CORP.  
 XX Takaiwa M, Okuda M, Saeki K, Kubota H, Hitomi J, Kageyama Y;  
 PI Shikata S, Nomura M;  
 XX P-PSDB; AAY17091.  
 DR WPI; 1999-287736/27.  
 DR P-PSDB; AAY17091.  
 XX  
 PT Alkali protease from Bacillus used in washing powders.  
 XX Disclosure; Page 63-68; 71pp; Japanese.  
 CC The invention relates to alkaline proteases produced by strains of  
 CC Bacillus. The proteases ability to digest casein is not inhibited by  
 CC oleic acid and they have a high stability to oxidising agents. The  
 CC alkaline protease of the invention has the following properties: (a) it  
 CC is active over the pH range 4-13 and has at least 80% of its optimum  
 CC activity over the range pH 6-12; (b) after 30 minutes at 40 deg. C it is  
 CC stable over the pH range 6-11; c) its isoelectric point is 8.9-9.1; (d)  
 CC its ability to digest casein is not inhibited by oleic acid; (e) it has  
 CC molecular weight about 43,000 by SDS-PAGE. The alkaline proteases can be  
 CC used as enzymes in washing compositions for use in automatic dishwashers  
 CC and for washing clothes. The stability to oxidising agents allows the  
 CC enzyme to be an effective component of washing compositions including  
 CC bleaches. The present sequence represents an alkaline protease encoding  
 CC DNA. (Updated on 20-MAR-2003 to correct DR field.)  
 XX  
 SQ Sequence 1923 BP; 578 A; 417 C; 474 G; 454 T; 0 U; 0 Other;

Alignment Scores:  
 Pred. No.: 8,698-160 Length: 1923  
 Score: 2120.50 Matches: 405  
 Percent Similarity: 97.70% Conservative: 19  
 Best Local Similarity: 93.32% Mismatches: 9  
 Query Match: 94.16% Indels: 1  
 DB: 2 Gaps: 1

US-09-985-689a-7 (1-433) x AAX37279 (1-1923)

QY 1 AsnAspValAlaArgGlyLeuValLysAlaAspValAlaGlnAsnAsnPhcGlyLeuTyr 20  
 DB 619 AATGATGTTGCGGTGAATGTTCAAGCGGATGTTGGCTCAGACGAGTACGGGTGAT 678  
 QY 21 GlyGlnGlyGlnLeuValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40  
 DB 679 GGACAGGACAGATGATACCGGTTCCGATACAGGCGTTGATACAGTTCGCAATGACAGT 738  
 QY 41 SerMetHisGluAlaPheArgGlyLysLeuThrAlaLeuTyrAlaLeuGlyArgThrAsn 60  
 DB 739 TCGATGATGAAGCTTCGCGGGAAATTAATGATGATGATGATGATGATGATGATGATGAT 798  
 QY 61 AsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn--- 79  
 DB 799 AATGCCAATGATACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 858  
 QY 80 AlaThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 99  
 DB 859 TCCACTAATAGGAATGGCGCTCAGGCGAATCTAGTCTTCAATCTATCATGATGATGATGAT 918  
 QY 100 GlyGlyGlyLeuGlyGlyLeuProAlaAsnLeuGlnThrLeuPheSerClnAlaTyrSer 119

Db 919 GGTGGGGACTTGGAGACTTACCTCGAATCTGCAACCTTATTCAGCAAGCATACAGT 978  
 QY 120 AlaGlyAlaArgIleHisThrAsnSerTrpGlyValaProValaAsnGlyAlaTyrThr 139  
 Db 979 GCTGTGCCAGATTCATACAACTCTCTGGGAGCAGCAGTGAATGGGGCTTACCAACA 1038  
 QY 140 AspSerArgAsnValaAspTyrValaArgLysAsnAspMetThrIleLeuPheAlaAla 159  
 Db 1039 GATTCAGAAATGTGGATGACTATGTGCGCAAAATGATATGACGATCCCTTTTGGTGGC 1098  
 QY 160 GlyAsnGluGlyProGlySerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 179  
 Db 1099 GGGAAATGAAGACCGGAAACCATCATGTCACCGACGACGATGAATATGCAATA 1158  
 QY 180 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 199  
 Db 1159 ACAGTCGGAGCTACGGAAACCTCGGCCAAGCTTTGGGTCTTATGCGGCAATATCAAC 1218  
 QY 200 HisValAlaGlnPheSerSerArgGlyProThrArgAspGlyArgIleLysProAspVal 219  
 Db 1219 CATGTGGCACAGTTCTCTTCACGTGGACCGCAAGAGTGGACGGATCAACACCGATGTC 1278  
 QY 220 MetAlaProGlyThrTyrIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 239  
 Db 1279 ATGGCACCGGAAACGTTCTATCTACTATCAGCAAGATCTTCTTGCACCGATTCCTCTTC 1338  
 QY 240 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProIle 259  
 Db 1339 TGGCGGAACCATGACAGTAAATATGATPACATGTTGGTGGAGCTCCATGGCTACACCGATC 1398  
 QY 260 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyValThrPro 279  
 Db 1399 GTTGTGTGAACGCTGGCAGCTTCGTGAGCATTTGTGAAAACAGAGGATCACACCA 1458  
 QY 280 LysProSerLeuLeuLysAlaAlaLeuLeuAlaGlyAlaAlaAspValGlyLeuGlyPhe 299  
 Db 1459 AAGCCTTCTCTATTAAAGCGGCACTGATTCGCGGTGCGAGTGCATCGGCTTGGCTAC 1518  
 QY 300 ProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaPhe 319  
 Db 1519 CCGAACGGTAACCAAGGATGGGAGCGATGACATTTGGATTAATCCCTGAAAGTTGCTTAT 1578  
 QY 320 ValAsnGluThrSerProLeuSerThrSerGlnLysAlaThrTyrSerPheThrAlaGln 339  
 Db 1579 GTGAAGAGTCCAGTCTCTATCCACAGCCAAAGGAGCGATCTCGTTTACTGCTACT 1638  
 QY 340 AlaGlyLysProLeuLysIleSerLeuValTrpSerAspAlaProGlySerThrThrAla 359  
 Db 1639 GCCGCAAGCCTTTGAAATCTCCCTGGTATGGTCTGATGCTGCGGCTGGAGGACCACTGCT 1698  
 QY 360 SerLeuThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrLysTyr 379  
 Db 1699 TCGGTAACTGTTGCAATGATCTGGACCTTGTCTATTCGCTCCAAATGGCACACAGTAT 1758  
 QY 380 ValGlyAsnAspPheThrAlaProTyrAspAsnAsnTrpAspGlyArgAsnAsnValGlu 399  
 Db 1759 GTAGGAATGACTTTACTTCGCCATACAAATGATACTGGGATGGCGCGCAATTAACGTAGAA 1818  
 QY 400 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrValGluValGlnAlaTyrAsn 419  
 Db 1819 AATGATATTTAATATGCACCAACAAACGGGAGCTATACATTAAGTACAGGCTTATAAC 1878  
 QY 420 ValProValSerProGlnThrPheSerLeuAlaIleValHis 433  
 Db 1879 GTACCGGTGGACACAGAACTTCTCGTTGGCAATTTGTGAAT 1920

## RESULT 4

AAX37277

ID AAX37277 standard; DNA; 1920 BP.

XX

AC AAX37277;

XX



AAQ27516  
 ID AAQ27516 standard; DNA; 1299 BP.  
 AC  
 XX AAQ27516;  
 DT 05-FEB-1993 (first entry)  
 XX  
 DE Alkali-protease Ya enzyme gene.  
 XX  
 KW Alkali resistance; surface active agent resistance; detergency improver;  
 KW ss.  
 XX  
 OS Bacillus sp. Y.  
 XX  
 XX Key Location/Qualifiers  
 FT 1. 1299  
 FT CDS /\*tag= a  
 XX  
 XX JF04197182-A.  
 PD 16-JUL-1992.  
 XX  
 XX 28-NOV-1990; 90JP-00327110.  
 XX  
 XX 28-NOV-1990; 90JP-00327110.  
 XX  
 XX (LLOY) LION CORP.  
 XX  
 XX WPI; 1992-288440/35.  
 DR P-PSDB; AAR26274.  
 XX  
 PT DNA coding alkali-protease Ya enzyme - has good alkali and surfactant  
 PT resistance and improves detergency.  
 XX  
 PS Claim 3; Page 2; 17pp; Japanese.  
 XX  
 CC The sequence is that of the alkali-protease Ya enzyme gene which can be  
 CC used in the recombinant production of Ya enzyme. Ya enzyme is excellent  
 CC in alkali resistance and surface active agent resistance and improves  
 CC detergency  
 XX  
 SQ Sequence 1299 BP; 401 A; 246 C; 297 G; 355 T; 0 U; 0 Other;

## Alignment Scores:

Pred. No.: 6.2e-153 Length: 1299  
 Score: 2032.00 Matches: 384  
 Percent Similarity: 94.46% Conservative: 25  
 Best Local Similarity: 88.68% Mismatches: 24  
 Query Match: 90.23% Indels: 0  
 DB: 2 Gaps: 0

US-09-985-689A-7 (1-433) x AAQ27516 (1-1299)

QY 1 AsnAspValAlaArgGlyLeuValAlaAspValAlaGlnAsnAsnPhsGlyLeuTyr 20  
 Db 1 AATGATGTAGCAAGAGGAGTAGTAAGCTGATGTTGCACAAACAAATACGGATTATAT 60  
 QY 21 GlyGlnGlyGlnLeuValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40  
 Db 61 GGACAAAGGTCAACTAGTTGCAGTAGCGGACACAGGCTTAGATACAGGTCGTAACGATAGT 120  
 QY 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60  
 Db 121 TCTATGATGAAGCAATTCGCGGGAATATCAGAGCTTTACGCGTTAGGAGAACTPAT 180  
 QY 61 AsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnAla 80  
 Db 181 AATGCGATGATCCGAATGGCATGGCAGACACATGTAGCAGGTTCTGTACTTTGGTAATGCT 240  
 QY 81 ThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSerGly 100  
 Db 241 TTTAAATAAGGAATGGCTCCGCAAGCTAACTTAGTCTTCCAAATCTATTATGATAGCAGC 300

## RESULT 6

AAAT85667  
 ID AAAT85667 standard; DNA; 1977 BP.  
 XX  
 AC AAAT85667;

QY 101 GlyGlyLeuGlyGlyLeuProAlaAsnLeuGlnThrLeuPheSerGlnAlaTyrSerAla 120  
 Db 301 GGAGGATTAGGTGGCTTACCATCGAACTTAAATACGTTATTATTAGTCAAGCTTGAATGCT 360  
 QY 121 GlyAlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyrThrThrAsp 140  
 Db 361 GGAGCAAGATTACATACTAATCTTTGGGAGCCCGCACTAAATGGAGCGTACACCTGCTAAC 420  
 QY 141 SerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAlaGly 160  
 Db 421 TCGAGACAAAGGGATGATGTTTCGAAATTAATGATATGACGGTACCTTTTTCGAGCTGGT 480  
 QY 161 AsnGluGlyProGlySerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThr 180  
 Db 481 AATGAGGTCCTAATTCAGGAACAATTAGTCTCCAGGTACAGCGAAAAATGCTATTACG 540  
 QY 181 ValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsnHis 200  
 Db 541 GTCGGCGCAACGGAAACTATCGCCCAAGCTTCGGTTCGATAGCAGATAACCCAAATCAT 600  
 QY 201 ValAlaGlnPheSerSerArgGlyProThrArgAspGlyArgIleLysProAspValMet 220  
 Db 601 AITGCACAAATTTTCATCGAGAGGAGCTACGAGGATGGACGAATTAAGCCTGACGTAACA 660  
 QY 221 AlaProGlyThrTyrIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrp 240  
 Db 661 GCTCCTGGAACATTTATTTATCAGCAGCTTCTCTTAGCTCCAGACTCTTCGTTTGG 720  
 QY 241 AlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIleVal 260  
 Db 721 GCGAATTTAACAGTAATACCGTATATGGCGGTACTCTCCATGGCGACACCTATTGTT 780  
 QY 261 AlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyValThrProLys 280  
 Db 781 GCAGGGAATGTCGCGCAATTACGTGAGCATTTTATAAAAAATAGAGGTATTACTCTCTAAG 840  
 QY 281 ProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspValGlyLeuGlyPhePro 300  
 Db 841 CCTTCTTTAAATAAAGCTGCATTTATCGTGTGCTACTGATGTTGGTTTAGGATATCT 900  
 QY 301 AsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaPheVal 320  
 Db 901 AGTGGTGACCAAGGCTGGGGGCGTGTACTCTAGATAAATCGTTAAATGTAGCGTATGTC 960  
 QY 321 AsnGluThrSerProLeuSerThrSerGlnLysAlaThrTyrSerPheThrAlaGlnAla 340  
 Db 961 AATGAAGCAACTGCATTAGCCACAGACAAAAAGCAAGCTATTCGTTTCCAGCACACAGG 1020  
 QY 341 GlyLysProLeuLysIleSerLeuValTrpSerAspAlaProGlySerThrThrAlaSer 360  
 Db 1021 GGTAAACCTTTAAATCTCGTTAGTAGCAGATGCTCTCTGGAAGTACAACTGCATCT 1080  
 QY 361 LeuThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrLysTyrVal 380  
 Db 1081 TATPACACTAGTTAATGATTTAGATCTAGTATTACTGCTCCGAATGGACAAAAATATGTA 1140  
 QY 381 GlyAsnAspPheThrAlaProTyrAspAsnAsnTrpAspGlyArgAsnAsnValGluAsn 400  
 Db 1141 GGAATGATTTAGTTATCTTATGATATACTGGATGGTGGCGCAACAATGTTGAGAAC 1200  
 QY 401 ValPheIleAsnAlaProGlnSerGlyThrThrValGluValGlnAlaTyrAsnVal 420  
 Db 1201 GTATTTTATAAACGCTCCGAATCTGGAACGTATATAATTGAGGTTCCAGCGTATATGTA 1260  
 QY 421 ProValSerProGlnThrPheSerLeuAlaIleValHis 433  
 Db 1261 CCATCTGGCCACAGCGTTCTCAGTATCGTACAT 1299

XX 17-OCT-2003 (revised)  
 DT 20-APR-1998 (first entry)  
 XX Thermococcus protease coding sequence.  
 XX Protease; research reagent; thermal stability; thermococcus celer; ss.  
 XX Thermococcus celer; DSM-2476.  
 OS WO9721823-A1.  
 PN 19-JUN-1997.  
 XX 07-NOV-1996; 96WO-JP003253.  
 XX 12-DEC-1995; 95JP-00323285.  
 XX (TAKI ) TAKARA SHUZO CO LTD.  
 XX Takakura H, Morishita M, Yamamoto K, Mitta M, Asada K;  
 PI Tsunashawa S, Kato I;  
 XX WPI; 1997-332794/30.  
 DR P-PSDB; AAW24121.  
 XX Protease(s) and genes encoding them obtained from Thermococcus and  
 PT Pyrococcus strains - have extremely high thermal stability and are useful  
 PT industrially and as research reagents.  
 XX Claim 3; Page 86-87; 159pp; Japanese.  
 CC This sequence represents the coding sequence for the protease from  
 CC Thermococcus celer DSM-2476. This sequence encodes a protease of the  
 CC invention. The proteases of the invention have extremely high thermal  
 CC stability. The proteases can be used as research reagents, and  
 CC industrially in the food, drug and chemical industries. (Updated on 17-  
 CC OCT-2003 to standardise OS field)  
 XX Sequence 1977 BP; 453 A; 659 C; 554 G; 311 T; 0 U; 0 Other;  
 SQ

Alignment Scores:  
 Pred. No.: 5,08e-25 Length: 1977  
 Score: 432.50 Matches: 134  
 Percent Similarity: 44.85% Conservative: 62  
 Best Local Similarity: 30.66% Mismatches: 156  
 Query Match: 19.21% Indels: 85  
 DB: 2 Gaps: 16

US-09-985-689a-7 (1-433) x AAT85667 (1-1977)

QY 8 ValtyeAlaAspValAlaGlnAsnAsnPhGlyLeuTyrglyGlnGlyGlnIleValAla 27  
 DB 433 ATAGGGCCGATACCGTCTGGAATCCCTCGGTACGAGGAGCGGTGGTGGTGGC 492  
 QY 28 ValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisGluAlaPheArg 47  
 DB 493 ATCGTCGATACGGGTATAGACGCGAAC-----CACCCCGATCTGAAG 534  
 QY 48 GlyLysIleThrAlaLeuTyrr---AlaLeuGlyArgThrAsnAsnAlaAsnAspProAsn 66  
 DB 535 GGCAGGTGTCATGGCTGATACGCGCGGTCTACGCGAGGTGCGACCCCTACGATGACCAG 594  
 QY 67 GlyHisGlyThrHisValAlaGlySerValLeuGly-----AsnAlaThrAsn 82  
 DB 595 GGACACGGAAACCCACGTTGGGGGTATCGTTGCCGGAACCGGACGCGTTAACTCCCAAGTAC 654  
 QY 83 LysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMet-----AspSerGly 100  
 DB 655 ATAGGGCTGCGCCCGCGGCGAAGCTCTCGGGTCAAGGTTCTCGGTGCGGACGCTTCG 714  
 QY 101 GlyGlyLeuGlyGlyLeuProAlaAsnLeuGlnThrLeuPheSerGlnAlaTyrrSerAla 120

Db 715 GGAAGCGTCTCCACCATCATCGCGGGTGTGACTGGTCTGCAGAACAGGACAAGTAC 774  
 QY 121 GlyAlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyrrThrAsp 140  
 Db 775 GGGATAAGGGTTCATCAACCTCTCCCTCGGCTCTCCCGAGAGCTCCGACGGAACCGACTCC 834  
 QY 141 SerArgAsnValAspAspTyrrValArgLysAsnAspMetThrIleLeuPheAlaAlaGly 160  
 Db 835 CTCAGTCAGGCGGTCAACAACGCGCTGGGACGCGGTATATAGTCTCGTCCGCGCCGGC 894  
 QY 161 AsnGluGlyProGlySerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThr 180  
 Db 895 AACAGCGGCGGAACACCTACCGTGGGTCTACCGCGCCGCCGAGCAAGGTCTATAACC 954  
 QY 181 ValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrrAlaAspAsnIleAsnHis 200  
 Db 955 GTCGGTGA-----GTTGACACCAACGACAAC 981  
 QY 201 ValAlaGlnPheSerSerArgGlyProThrArgAspGlyArgIleLysProAspValMet 220  
 Db 982 ATCGCCAGCTTCTCCAGCAGGGGACCGACCGCGGACGGAAGGTCTCAAGCCGGAAGTCGTC 1041  
 QY 221 AlaProGlyThrTyrrIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrp 240  
 Db 1042 GCCCGCGCGTTCATCATAGCCCGCGCGCCAGC-----GGAACGACATCGGC 1092  
 QY 241 AlaAsnHisAspSerLysTyrrAlaTyrrMetGlyThrSerMetAlaThrProIleVal 260  
 Db 1093 ACCCGATAACGACTACTACCAAGCGCTCTGGACACGATGCGCACCCGACGTT 1152  
 QY 261 AlaGly---AsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyValThrPro 279  
 Db 1153 TCGGCGGTGGCGCGCTCATCTCCAGCGCCAC-----CCG 1188  
 QY 280 Lys-----ProSerLeuLeuLysAlaAlaIleAlaGlyAla----- 292  
 Db 1189 AGTGGACCCCGGACAGGGAAGACCGCTCTTCAGACCGCCGACATAGTCGCGCCC 1248  
 QY 293 -----AlaAspValGlyLeuGlyPheProAsnGlyAsnGlnGlyTrpGlyArgVal 309  
 Db 1249 AAGGAGATAGCGGACATCGCTACGGTGC-----GGTAGGGTG 1287  
 QY 310 ThrLeuAspLysSerLeu-----AsnValAlaPheValAsnGluThrSerProLeu 326  
 Db 1288 AACGTCTACAAGGCCATCAAGTACGACGACTACGCCAAGCTCACCTTCACCGGCTCCGTC 1347  
 QY 327 SerThrSerGlnLysAlaThrTyrrSerPheThrAlaGlnAlaGlyLysProLeuLysIle 346  
 Db 1348 GCCGACAAGGGAAGCGCCACCCACCTTC----- 1377  
 QY 347 SerLeuValTrpSerAspAlaProGlySerThrThrAlaSerLeuThrLeuVal----- 364  
 Db 1378 -----GACGTACGCGCGCCACCTTCGTGACCGCCACCTCTACTGGGAC 1422  
 QY 365 -----AsnAspLeuAspLeuValIleThrAlaProAsnGlyThrLysTyrrValGly 381  
 Db 1423 ACGGCTCGACGACATCGACTCTACTCTACACCCCAACGGAAGAG---GTTGAC 1479  
 QY 382 AsnAspPheThrAlaProTyrrAspAsnTrpAspGlyArgAsnAsnValGluAsnVal 401  
 Db 1480 TACTCTACACCGCTACTAC-----GGCTTCGAGAAGGTC 1515  
 QY 402 PheIleAsnAlaProGlnSerGlyThrTyrrValGluValGlnAlaTyrr 418  
 Db 1516 GGCTACTACAACCGACCGCGGAACCTGGACGCTCAAGTCTCAGTCTCAGCTAC 1566

## RESULT 7

AA05926

ID AAX05926 standard; DNA; 1977 BP.

XX AAX05926;

XX 06-MAY-1999 (first entry)

DT



XX W09856926 Seq ID 11.  
 XX Hyperthermostable; protease; thermophilic; bacterium; subtilisin;  
 KW additive; drug; washing agent; foodstuff; chemical synthesis; ds.  
 XX Thermococcus celer.  
 OS Thermococcus celer.  
 XX W09856926-A1.  
 XX 17-DEC-1998.  
 XX 04-JUN-1998; 98WO-JP002465.  
 XX 10-JUN-1997; 97JP-00151969.  
 XX (TAKI ) TAKARA SHUZO CO LTD.  
 XX Takakura H, Morishita M, Shimojo T, Asada K, Kato I;  
 XX WPI; 1999-080907/07.  
 XX P-PSDB; AAW94840.  
 XX Recombinant hyperthermostable protease from *Pyrococcus furiosus* - and  
 PT gene encoding it, for large scale production of the protease for  
 PT industrial use.  
 XX Disclosure; Page 53-54; 82pp; Japanese.  
 XX The invention relates to a hyperthermostable protease derived from a  
 CC thermophilic bacterium (especially *Pyrococcus furiosus*). The protease has  
 CC working temperature 40-110 deg.C (optimum 90-95 deg.C), working pH 5-10  
 CC (optimum 6-8), and retains more than 90% of its activity after 8 hours at  
 CC 95 deg.C. The invention also provides gene sequences encoding a  
 CC polypeptide of formula SIG-Ala-Gly-Asn-PRO, where SIG is a signal  
 CC peptide from subtilisin, and PRO is the above protease. Host cells  
 CC (especially *Bacillus* strains) transformed with vectors comprising the  
 CC genes are used for the recombinant production of the protease. The  
 CC hyperthermostable protease which can be prepared in quantity suitable for  
 CC industrial use, can be used as an additive for drugs, washing agents and  
 CC foodstuffs and for chemical synthesis  
 XX SQ Sequence 1977 BP; 453 A; 659 C; 554 G; 311 T; 0 U; 0 Other;

Alignment Scores:  
 Pred. No.: 5.08e-25 Length: 1977  
 Score: 432.50 Matches: 134  
 Percent Similarity: 44.85% Conservative: 62  
 Best Local Similarity: 30.66% Mismatches: 156  
 Query Match: 19.21% Indels: 85  
 DB: 2 Gaps: 16

US-09-985-689A-7 (1-433) x AAX05926 (1-1977)

QY 8 VallysAlaAspValAlaGlnAsnAsnPheGlyLeuTyGlyGlnIleValAla 27  
 DB 433 ATAGGGCGGATACCGTCTGGAACTCCCTCGCTACGACGAGGCGGTGGTGGTGC 492  
 QY 28 ValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerMetHisGluAlaPheArg 47  
 DB 493 ATCTGCGATACGGGTATAGACCGGAAC-----CACCCGATCTCGAAG 534  
 QY 48 GlyLysIleThrAlaLeuTyf---AlaLeuGlyArgThrAsnAlaAsnAspProAsn 66  
 DB 535 GGCNAGGTCATAGGTGGTGTACGACCGCTCAACGCGAGGTGCGACCCCTACGATGACCAG 594  
 QY 67 GlyHisGlyThrHisValAlaGlySerValLeuGly-----AsnAlaThrAsn 82  
 DB 595 GGACACGGGAACCCAGTTCGGGTATCGTTGCCGAACCGGACGCGGTTAACTCCGAGPAC 654  
 QY 83 LysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMet-----AspSerGly 100  
 DB 655 ATAGGGCTCGCCCCCGGCGAAGCTCTGCGGCTCAAGGTCTCGGTGCGGACGGTTCG 714

QY 101 GlyGlyLeuGlyGlyLeuProAlaAsnLeuGlnThrLeuPheSerGlnAlaTySerAla 120  
 DB 715 GGAAGCGTCTCCACCATCATCGCGGTGTTGACTGGGTCTCCAGACACAGGCAAGTAC 774  
 QY 121 GlyAlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyThrThrAsp 140  
 DB 775 GGGATAAGGGTCATCAACCTCTCTCGGTCTCTCCAGAGCTCCGACGGAACGACTCC 834  
 QY 141 SerArgAsnValAspAspTyfValArgLysAsnAspMetThrIleLeuPheAlaAlaGly 160  
 DB 835 CTCAGTCAGCGCTCAACACGCGCTGGACCGCGTATAGTAGTCTCGTCCGCCCGGC 894  
 QY 161 AsnGluGlyProGlySerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThr 180  
 DB 895 AACAGCGGGCGAACAACCTACACGCTCGGTCTACCCCGCGCGAGCAAGTCTATAACC 954  
 QY 181 ValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyfAlaAspAsnIleAsnHis 200  
 DB 955 GTCGGTGCA-----GTTGACGCAACGACAAC 981  
 QY 201 ValAlaGlnPheSerSerArgGlyProThrArgAspGlyArgIleLysProAspValMet 220  
 DB 982 ATCGCCAGCTTCTCCAGCAGGGACCGACCGCGAGCGAAGGCTCAAGCCGGAAGTCTC 1041  
 QY 221 AlaProGlyThrTyfIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrp 240  
 DB 1042 GCCCGCGGTGACATCATAGCCCGCGCGCCAGC-----GGMACGAGCATGGGC 1092  
 QY 241 AlaAsnHisAspSerLysTyfAlaTyfMetGlyGlyThrSerMetAlaThrProIleVal 260  
 DB 1093 ACCCGATAACGACTACTACACCAAGCGCTCTGACACGACATGGCCACCGCAGCTT 1152  
 QY 261 AlaGly---AsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyValThrPro 279  
 DB 1153 TCGGGCGTGGCGCTCATCTCCAGCCCGCAC-----CCG 1188  
 QY 280 Lys-----ProSerLeuLysAlaAlaLeuIleAlaGlyAla----- 292  
 DB 1189 AGCTGGACCCCGGACGAGTGAAGACCGCTCATCGAGACCGCGACATAGTCGCCCC 1248  
 QY 293 -----AlaAspValGlyLeuGlyPheProAsnGlyAsnGlnGlyTyfTrpGlyArgVal 309  
 DB 1249 AAGGAGATAGCGGACATCGCTACGTCGCG-----GGTAGGGTG 1287  
 QY 310 ThrLeuAspLysSerLeu-----AsnValAlaPheValAsnGluThrSerProLeu 326  
 DB 1288 AACGTCATACAGGCCATCAAGTACGACGACTACGCAAGCTCACCTTACCGGCTCCGTC 1347  
 QY 327 SerThrSerGlnLysAlaThrTyfSerPheThrAlaGlnAlaGlyLysProLeuLysIle 346  
 DB 1348 GCGACACGAGGGAAGCGCCACCCACACCTTC----- 1377  
 QY 347 SerLeuValTrpSerAspAlaProGlySerThrThrAlaSerLeuThrLeuVal----- 364  
 DB 1378 -----GACGTACGCGCGCCACCTTCGTGACCGCCACCCCTCTACTGGGAC 1422  
 QY 365 -----AsnAspLeuAspLeuValIleThrAlaProAsnGlyThrLysTyfValGly 381  
 DB 1423 ACGGGCTCGAGGACATCGACCTTACCTCTACGACCCCAACGGAACGAG---GTTGAC 1479  
 QY 382 AsnAspPheThrAlaProTyfArgAsnAsnTrpAspGlyArgAsnAsnValGluAsnVal 401  
 DB 1480 TACTCTACACCGCTTACTAC-----GGCTTCGAGAGGTC 1515  
 QY 402 PheIleAsnAlaProGlnSerGlyThrTyfThrValGluValGlnAlaTyf 418  
 DB 1516 GGCTACTACACCGGACCGCGAAGCTGAGCGGTCAAGGTCTCGTCACTAC 1566

RESULT 8  
 AAX05920  
 ID AAX05920 standard; DNA; 1236 BP.  
 XX



AA085668  
ID AA085668 standard; DNA; 1566 BP.  
XX  
AC AA085668;  
XX  
DT 17-OCT-2003 (revised)  
DT 20-APR-1998 (first entry)  
XX  
XX Pyrococcus furiosus protease coding sequence.  
DE  
XX Protease; research reagent; thermal stability; pyrococcus furiosus; ss.  
XX  
XX Pyrococcus furiosus; DSM-3638.  
XX  
XX Pyrococcus furiosus; DSM-3638.  
XX  
XX Key Location/Qualifiers  
FT CDS 1..1566  
FT /tag= a  
FT /transl\_except= (pos: 1282..1284, aa: Xaa)  
FT /note= "Xaa= Gly, Val"  
XX  
XX WO9721823-A1.  
XX  
XX PD 19-JUN-1997.  
XX  
XX PF 07-NOV-1996; 96WO-JP003253.  
XX  
XX PR 12-DEC-1995; 95JP-00323285.  
XX  
XX PA (TAKI) TAKARA SHUZO CO LTD.  
XX  
XX PI Takakura H, Morishita M, Yamamoto K, Mitta M, Asada K;  
PI Tsunashawa S, Kato I;  
DR WPI; 1997-332794/30.  
DR P-PSDB; AAW24122.  
XX  
XX Protease(s) and genes encoding them obtained from Thermococcus and  
PT Pyrococcus strains - have extremely high thermal stability and are useful  
PT industrially and as research reagents.  
XX  
XX Claim 7; Page 90-91; 159pp; Japanese.  
XX  
XX This sequence represents the coding sequence for the protease from  
CC Pyrococcus furiosus DSM-3638. This sequence encodes a protease of the  
CC invention. The proteases of the invention have extremely high thermal  
CC stability. The proteases can be used as research reagents, and  
CC industrially in the food, drug and chemical industries. (Updated on 17-  
CC OCT-2003 to standardise OS field)  
XX  
XX SQ Sequence 1566 BP; 467 A; 342 C; 372 G; 384 T; 0 U; 1 Other;  
  
Alignment Scores:  
Pred. No.: 3,12e-22 Length: 1566  
Score: 396.00 Matches: 127  
Percent Similarity: 42.56% Conservative: 56  
Best Local Similarity: 29.53% Mismatches: 163  
Query Match: 17.58% Indels: 84  
DB: 2 Gaps: 15  
  
US-09-985-689A-7 (1-433) x AA085668 (1-1566)  
QY 16 AsnPhGlyLeuTyrglyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThr 35  
DB 58 AACTTGGATATGATGTTCTGGAATCAACATAGGAATTAATGACACTGGAATTGAC--- 114  
QY 36 GlyArgAsnAspSerSerMetHisGluAlaPheArgGlyIleThrAlaLeuTyra 55  
DB 115 -----GCTTCTCATCCAGATCTCCAGGAAGAAGTA----- 144  
QY 56 LeuGlyArgThrAsnAsnAlaAsn-----AspProAsnGlyHisGly 69  
DB 145 ATTGGGTGGGTAGATTGTCATGTAGGAGTTATCCATGATGACCATGACATGACATGGA 204

QY 70 ThrHisValAlaGlySerValLeuGly-----AsnAlaThrAsnLysGly 84  
DB 205 ACTCATGTAGCTTCAATAGCAGCTGGTACTGGAGCAGCAAGTAATGGCAAGTACAGGGA 264  
QY 85 MetAlaProGlnAlaAsnLeuValPheGlnSerIleMet-----AspSerGlyGlyGly 102  
DB 265 ATGGCTCCAGGAGCTAAGCTGGCGGAATTAAGGTTCTAGGTGCGGATGGTCTTGGAGC 324  
QY 103 LeuGlyGlyLeuProAlaAsnLeuGlnThrLeuPheSerGlnAlaTySerAlaGlyAla 122  
DB 325 ATATCTACTATAATTAAGGGAGTTGAGTGGCCGTTGATACAAAGATAAGTACGGAATT 384  
QY 123 ArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyThrThrAspSerArg 142  
DB 385 AAGGTCAATTAATCTTTCTTGGTTCAAGCCAGCAGCTCAGATGGTACTGACGCTCTAAGT 444  
QY 143 AsnValAspAspTyValArgLysAsnAspMetThrIleLeuPheAlaAlaGlyAsnGlu 162  
DB 445 CAGGCTGTTAATGACGCGTGGATCTCGATTAGTTGTTGTTGGTTGCGCTGGAAACAGT 504  
QY 163 GlyProGlySerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrValGly 182  
DB 505 GGACCTAACACAGTATACATCGTTCTCCAGCAGCTGCAAGCAAAAGTTATTACAGTTGGA 564  
QY 183 AlaThrGluAsnLeuArgProSerPheGlySerTyAlaAspAsnIleAsnHisValAla 202  
DB 565 GCC-----GTTGACAGTATGATGTTATAACA 591  
QY 203 GlnPheSerSerArgGlyProThrArgAspGlyArgIleLeuPheAspValMetAlaPro 222  
DB 592 AGCTTCTCAAGCAGAGGGCCAACTGCAGAGCGCAGCTTAAGCCTGAGGTTGTTGCTCCA 651  
QY 223 GlyThrTyrlleuSerAlaArgSerSerLeuAlaProAspSerSerPheTrpAlaAsn 242  
DB 652 GGAACCTGGATAATTGCTCCAGAGCAAGT-----GGAACTAGCATGGTCAACCA 702  
QY 243 HisAspSerLysTyAlaTyMetGlyGlyThrSerMetAlaThrProIleValAlaGly 262  
DB 703 ATTAATGACTATTACACAGCAGCTCTGGGACATCAATGGCACTCTCACGTAGCTGT 762  
QY 263 AsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyValThrProLys----- 280  
DB 763 ATTGCAGCCCTCTTGTCTCAA-----GCACACCCGAGCTGGACT 801  
QY 281 ProSerLeuLysAlaAlaLeuIleAlaGlyAla----- 292  
DB 802 CCAGACAAAGTAAACAGCCCTCATAGAACTGTCTGATATCGTAAGCCAGATGAAATA 861  
QY 293 AlaAspValGlyLeuGlyPheProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAsp 312  
DB 862 GCGGATATACCTACCGTGC-----GGTAGGTTAATGATATAC 900  
QY 313 LysSerLeu-----AsnValAlaPheValAsnGluThrSerProLeuSerThrSer 329  
DB 901 AAGCTATAAATACGATACATATGCAAGCTAGTGTCTCAGTATGTTGCCAACA 960  
QY 330 GlnLysAlaThrTySerPheThrAlaGlnAlaGlyLysProLeuLysIleSerLeuVal 349  
DB 961 GGCAGCCAAACTCACCAGTTCTGTTATTAGCGAGCTCGTTCGTATGTCACCATATATAC 1020  
QY 350 TrpSerAspAlaProGlySerThrThrAlaSerLeuThrLeuValAsnAspLeuAspLeu 369  
DB 1021 TGGACAAATGCCAAT-----AGCGACCTTGTATCTT 1050  
QY 370 ValIleThrAlaProAsnGlyThrLysTyValGlyAsnAspPheThrAlaProTyThrAsp 389  
DB 1051 TACCTCTACGATCCCAATGGAACCCAG---GTTGACTACTCTTACACCGCTACTAT--- 1104  
QY 390 AsnAsnTrpAspGlyArgAsnAsnValGluAsnValPheIleAsnAlaProGlnSerGly 409  
DB 1105 -----GGATTCGAAAAGTGGTTATTACACACCACTGATGGA 1143  
QY 410 ThrTyThrValGluValGlnAlaTyAsn 419



## RESULT 11

AA05929  
ID AAX05929 standard; DNA; 1962 BP.  
AC AAX05929;  
XX  
DT 06-MAY-1999 (first entry)  
XX  
XX Hyperthermostable protease encoding DNA.  
DE Hyperthermostable; protease; thermophilic; bacterium; subtilisin;  
KW additive; drug; washing agent; foodstuff; chemical synthesis; ds.  
XX Pyrococcus furiosus.  
OS  
XX  
PN MO9856926-A.  
PD 17-DEC-1998.  
XX  
XX 04-JUN-1998; 98MO-JP002465.  
XX  
XX 10-JUN-1997; 97JP-00151969.  
XX  
XX (TAKI) TAKARA SHUZO CO LTD.  
XX  
XX Takakura H, Morishita M, Shimojo T, Asada K, Kato I;  
XX WPI; 1999-080907/07.  
XX P-PSDB; AAW94841.  
XX  
XX Recombinant hyperthermostable protease from Pyrococcus furiosus - and  
PT gene encoding it, for large scale production of the protease for  
XX industrial use.  
XX  
XX Disclosure; Page 59-60; 82pp; Japanese.  
PS  
XX The invention relates to a hyperthermostable protease derived from a  
CC thermophilic bacterium (especially Pyrococcus furiosus). The protease has  
CC working temperature 40-110 deg.C (optimum 80-95 deg.C), working pH 5-10  
CC (optimum 6-8), and retains more than 90% of its activity after 8 hours at  
CC 95 deg.C. The invention also provides gene sequences encoding a  
CC polypeptide of formula SIG-Ala-Gly-Ala-PRO, where SIG is a signal  
CC peptide from subtilisin, and PRO is the above protease. Host cells  
CC (especially Bacillus strains) transformed with vectors comprising the  
CC genes are used for the recombinant production of the protease. The  
CC hyperthermostable protease which can be prepared in quantity suitable for  
CC industrial use, can be used as an additive for drugs, washing agents and  
CC foodstuffs and for chemical synthesis  
XX  
SQ Sequence 1962 BP; 602 A; 399 C; 471 G; 490 T; 0 U; 0 Other;

## Alignment Scores:

Pred. No.: 4,13e-22 Length: 1962  
Score: 396.00 Matches: 127  
Percent Similarity: 42.56% Conservative: 56  
Best Local Similarity: 29.53% Mismatches: 163  
Query Match: 17.58% Indels: 84  
DB: 2 Gaps: 15

US-09-985-689A-7 (1-433) x AAX05929 (1-1962)

QY 16 AsnPheGlyLeuTyrglyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThr 35  
DB 454 AACTGGGATATGATGGTCTCGAATCAATAGGAAATATGGACACCTGGAAATGAC--- 510  
QY 36 GlyArgAsnAspSerSerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyra 55  
DB 511 -----GCTTCTCATCCAGATCTCCAGGAAAGTA----- 540  
QY 56 LeuGlyArgThrAsnAsnAlaen-----AspProAsnGlyHisGly 69  
DB 541 ATTGGGTGGGTAGATTGTTGTCATGTGTAGGAGTTATCCATAGCATGGACCATGGATGGA 600

QY 70 ThrHisValAlaGlySerValLeuGly-----AsnAlaThrAsnLysGly 84  
DB 601 ACTCATGTAGCTTCAATAGCAGCTGGTACTGGACGACCAAGTATATGCAAGTACAGGGA 660  
QY 85 MetAlaProGlnAlaAsnLeuValPheGlnSerIleMet-----AspSerGlyGlyGly 102  
DB 661 ATGGCTCCAGGAGCTAAGCTGGCGGAATTAAGGTCTTAGGTGCGCATGTTCTGGAGC 720  
QY 103 LeuGlyGlyLeuProAlaAsnLeuGlnThrLeuPheSerGlnAlaTyrsrAlaGlyAla 122  
DB 721 ATATCTACTATAATTAAGGGAGTTGAGTGGCGCTTGTATTAACAAGATAAGTACGGAAT 780  
QY 123 ArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyrsrThrAspSerArg 142  
DB 781 AAGGTCAATTAATCTTCTTCTTGGTTCAGCCAGAGCTCAGATGGTACTAGCGCTCTAAGT 840  
QY 143 AsnValAspAspTyrsrValArgLysAsnAspMetThrIleLeuPheAlaAlaGlyAsnGlu 162  
DB 841 CAGGCTGTTAATGCAGCGTGGGATGCTGGATTAGTTGTTGTTGCTGGCGTGGAAACAGT 900  
QY 163 GlyProGlySerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrValGly 182  
DB 901 GGACCTAACAAAGTATACAAATCGTTCCTCCAGCAGCTGCAAGCAAAAGTTATTACAGTTGA 960  
QY 183 AlaThrGluAsnLeuArgProSerPheGlySerTyrsrAlaAspAsnIleAsnHisValAla 202  
DB 961 GCC-----GTTGCAAGTATGATTATTATAACA 987  
QY 203 GlnPheSerSerArgGlyProThrArgAspGlyArgIleLysProAspValMetAlaPro 222  
DB 988 AGCTTCTCAAGCAGAGGGCCCACTGCAGCGGAGGCTTAAGCTGAGGTGTGTGCTCCA 1047  
QY 223 GlyThrTrpIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrpAlaAsn 242  
DB 1048 GAAACTCGATAATGCTGCCAGAGCAAGT-----GGAACACATGATGTTTATAACA 1098  
QY 243 HisAspSerLysTyrsrAlaTyrsrMetGlyThrSerMetAlaThrProIleValAlaGly 262  
DB 1099 ATTAATGACTATTACACAGCAGCTCTCGGACATCAATGGCACTCTCTCAGTAGCTGT 1158  
QY 263 AsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyValThrProLys----- 280  
DB 1159 ATTGAGCGCTCTTCTCTCAA-----GCACACCCGAGCTGGACT 1197  
QY 281 ProSerLeuLeuLysAlaAlaLeuIleAlaGlyAla----- 292  
DB 1198 CAGACAAAGTAAACACAGCCCTCATAGAACTCTCATATCGTAAGCCAGATGAATA 1257  
QY 293 AlaAspValGlyLeuGlyPheProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAsp 312  
DB 1258 GCCGATATAGCTTACGGTGCA-----GCTAGCGTTAATGCATAC 1296  
QY 313 LysSerLeu-----AsnValAlaPheValAsnGlnThrSerProLeuSerThrSer 329  
DB 1297 AAGGCTATAAAGTACGATATATGCAAGCTATGTTCTAGTATGTTGTCACAAACA 1356  
QY 330 GlnLysAlaThrTyrsrPheThrAlaGlnAlaGlyLysProLysIleSerLeuVal 349  
DB 1357 GGCAGCCAAACTCACCACTCGTTATTAGCGGAGCTTCGTTGCTGTAACATGCACATTATAC 1416  
QY 350 TrpSerAspAlaProGlySerThrThrAlaSerLeuThrLeuValAsnAspLeuAspLeu 369  
DB 1417 TGGGCAATGCCAAT-----ACGACCTTGTATCTT 1446  
QY 370 ValIleThrAlaProAsnGlyThrLysTyrsrValGlyAsnAspPheThrAlaProTyrsrAsp 389  
DB 1447 TACCTCTACGATCCCAATGGAAACAG---GTTGACTACTCTTACACCGCTACTAT--- 1500  
QY 390 AsnAsnTrpAspGlyArgAsnValGluAsnValPheIleAsnAlaProGlnSerGly 409  
DB 1501 -----GGATTCCGAAAGGTGTTATTATTACAAACCCCACTGATGGA 1539



ABL54900  
 ID ABL54900 standard; DNA; 2121 BP.  
 XX AC ABL54900;  
 XX DT 11-SEP-2003 (revised)  
 XX DT 31-MAY-2002 (first entry)  
 XX T. yonsei subtilisin-like serine protease coding sequence.  
 DE  
 XX  
 XX KW Subtilisin-like serine protease; ss.  
 XX OS Thermoanaerobacter yonseiensis.  
 XX Key Location/Qualifiers  
 XX FH 142..1779  
 XX CDS /\*tag= a  
 XX FT /product= "subtilisin-like serine protease"  
 XX FT  
 XX PN KR2000072141-A.  
 XX  
 XX PD 05-DEC-2000.  
 XX PF 04-AUG-2000; 2000KR-00045411.  
 XX PR 04-AUG-2000; 2000KR-00045411.  
 XX PA (KIMY/) KIM Y S.  
 XX FI Chang HJ, Kim DH, Byun YR, Kim YS;  
 XX DR WPI; 2001-298092/31.  
 XX DR P-PSDB; ABB09483.  
 XX New DNA sequence of thermophilic protein decomposition enzyme and protein  
 XX PT derived therefrom.  
 XX PS Claim 1; Page 6; 15pp; Korean.  
 XX CC This sequence represents the DNA encoding the Thermoanaerobacter yonsei  
 XX CC subtilisin-like serine protease of the invention. (Updated on 11-SEP-2003  
 XX CC to standardise OS field)  
 XX SQ Sequence 2121 BP; 712 A; 410 C; 425 G; 554 T; 0 U; 20 Other;  
 Alignment Scores:  
 Pred. No.: 5,93e-20 Length: 2121  
 Score: 369.50 Matches: 118  
 Percent Similarity: 48.73% Conservative: 57  
 Best Local Similarity: 31.55% Mismatches: 136  
 Query Match: 16.41% Indels: 63  
 DB: 4 Gaps: 14  
 US-09-985-689A-7 (1-433) x ABL54900 (1-2121)  
 QY 6 GlyIleValIysAlaAspValAlaGlnAsnAsnPheGlyLeuTyrglyGlnGlyGlnle 25  
 DB 445 GGAATCACAAA-----GCACGAGTGATTGGAGTCACAGGAAAATAATACA 495  
 QY 26 ValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisGluAla 45  
 DB 496 ATAGCAATATTGACACAGGTATACAGGAAATCACTGACCTCTCA----- 543  
 QY 46 PheArgGlyIleValIleThrAlaLeuTyrglyArgThrAsnAsnAlaAsn----- 63  
 DB 544 ---GGTGGAAAATA-----ATAGGATGGAAGACCTTTATCAACACAAA 585  
 QY 64 -----AspProAsnGlyHisGlyThrHisVal-----AlaGlySer 75  
 DB 586 ACTACACCATACGACCAATGCCATGCCATGCACTACGTAGCAAGTATTGCTGCAGGTACA 645  
 QY 76 ValLeuGlyAsnAlaThrAsnIysGlyMetAlaProGlnAlaAsnLeuValPheGlnSer 95

Db 646 GGTGCTGGAACAGTCTTTACAAAGCGTTGCTCTGATGCTTTGTTGGTAGGAATAAAA 705  
 QY 96 IleMetAspSer-----GlyGlyGlyLeuProAlaAsnLeuGlnThrLeu 113  
 DB 706 GTTTTAGATCAAAATGGAAGCGCAGCATGAGCACTGTAAGTCTGAGCAAGATTGCTGGCT 765  
 QY 114 PheSerGlnAlaTyrsrAlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaProVal 133  
 DB 766 GTTCAAAATAAAGATGTATACGGAATCAAAAGTTATAAATTTAAAGCTTCGCGACTTCACA 825  
 QY 134 AsnGlyAlaTyrrThrAspSerArgAsnValAspAspTyrrValArgLysAsnAspMet 153  
 DB 826 AGTTCTGATGCACTGACTCTACCTCATTCAGCATGATAGACGATAGTAGCGGTATT 885  
 QY 154 ThrIleLeuPheAlaAlaGlyAsnGluGlyProGlySerGlyThrIleSerAlaProGly 173  
 DB 886 GTAGTAGTTGTGACGAGCAAGAACTCTGGCCCTGCAAAATACACCATAGGTCCTCCCTGGT 945  
 QY 174 ThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsnLeuArgProSerPheGlySer 193  
 DB 946 GCTGCGGAAAAGCCATAACACAGTCGACCAATGCGACATGTA-----GGTGAA 993  
 QY 194 TyrAlaAspAsnIleAsnHisValAlaGlnPheSerSerArgGlyProThrArgAspGly 213  
 DB 994 CTGGCTTTAACCTT-----GCAAGCTTTCCAGCGCGTCTCTACTGCTGACGGA 1044  
 QY 214 ArgIleLysProAspValMetAlaProGlyThrTyrrIleLeuSerAlaArgSerSerLeu 233  
 DB 1045 AGATAAAACCTGACATTGCGGCCCCAGGATATAATATACTGCCGGAAG----- 1095  
 QY 234 AlaProAspSerSerPheTrpAlaAsnHisAspSerLysTyrrAlaTyrrMetGlyGlyThr 253  
 DB 1096 -----GCGAATTCTGTAATGGATATGTAACATACAGCGGTACA 1134  
 QY 254 SerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLys 273  
 DB 1135 AGCATGGCAACACCTTTTGTAGCAGAACTGTTGCTCTTATGCTTAAC-----GCT 1185  
 QY 274 AsnArgGlyValThrProLysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAla 293  
 DB 1186 AATCCAAATCTCCTCCAAATGATGCA-----AAAAATATAATATGCTACTGCAAAA 1239  
 QY 294 AspValGlyLeuGlyPheProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLys 313  
 DB 1240 AGCTGGGGGCTCCAAAGCAAAACGTTGACTACTGTGTCAGGAAAGATTGGATGGTATGAA 1299  
 QY 314 SerLeuAsnValAla-----Phe 319  
 DB 1300 GCTATTAAGGCTAGCAGGTAATTTTAGAGGAAATAATATTGATGTACCAATCATTTAT 1359  
 QY 320 ValAsnGluThrSerProLeuSerThrSerGlnLysAlaThrTyrrSerPheThrAla--- 338  
 DB 1360 ATTTCCAGGTATCTGCC-----GGTCTCGTTACAGTGATCTGGACTTTTAACGCAACA 1416  
 QY 339 GlnAlaGlyLysProLeuLysIleSerLeuValTrpSerAsp 352  
 DB 1417 AATACAAGTTATCCCAATCGCAATAACATTAATAATTCCTGAC 1458  
 RESULT 14  
 ID AAT61454 standard; DNA; 2539 BP.  
 XX AAT61454;  
 XX AC AAT61454;  
 XX 06-OCT-1997 (first entry)  
 XX Streptomyces viridosporus dhpa gene.  
 XX asymmetric hydrolase; dhpa; 4-substituted-1,4-dihydropyridine;  
 KW derivative; Streptomyces viridosporus; ester; chiral; synthesis;  
 XX cardiovascular; treatment; hypertension; ischaemic heart disease; ds.  
 XX Streptomyces viridosporus.  
 OS



```

XX FH Key Location/Qualifiers
XX CDS 338..2539
XX FT /*tag= a
XX FT /note= "no stop codon given"
XX FT misc_feature 950..2509
XX FT /*tag= b
XX FT /note= "encodes AAW13666"
XX PN WO9705243-A1.
XX XX 13-FEB-1997.
XX PF 30-JUL-1996; 96WO-JP002147.
XX XX 31-JUL-1995; 95JP-00212975.
XX PR 29-FEB-1996; 96JP-00067478.
XX XX
XX PA (SAOC ) MERCIAN CORP.
XX PI Arisawa A, Matsufuji M, Tsuruta T, Dobashi K, Nakaehima T;
XX PI Ishiki K, Yoshio K T;
XX XX
XX DR WP1; 1997-145682/13.
XX DR P-PSDB; AAW13666, AAW13667.
XX XX
XX PT Asymmetric hydrolase gene derived from Streptomyces viridosporus - acts
XX PT on 4-substituted-1,4-di:hydro:pyridine derivatives to produce chiral
XX PT derivatives useful for synthesis of cardiovascular drugs.
XX XX
XX PS Claim 3; Page 49-55; 78pp; Japanese.
XX XX
XX CC This sequence is the Streptomyces viridosporus dbpA gene which encodes an
XX CC asymmetric hydrolase which acts on 4-substituted-1,4-dihydropyridine
XX CC derivatives. The enzyme allows the efficient conversion of 4-substituted-
XX CC 1,4-dihydropyridine esters to chiral partially hydrolysed derivatives,
XX CC for use in the synthesis of cardiovascular drugs suitable for the
XX CC treatment of e.g. hypertension and ischaemic heart disease
XX XX
XX SQ Sequence 2539 BP; 433 A; 959 C; 867 G; 280 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: Length: 2539
Score: 327.50 Matches: 136
Percent Similarity: 40.83% Conservative: 51
Best Local Similarity: 29.69% Mismatches: 160
Query Match: 14.54% Indels: 111
DB: 2 Gaps: 23

US-09-985-689A-7 (1-433) x AAT61454 (1-2539)
Qy 18 GlyLeuTyrGlyGlnGlyGlnLeValAlaValAlaAspThrGlyLeuAspThrGlyArg 37
Db 998 GGCTACGACGGCAAGGGGTGAAGATCGCCGCTCTGGACACCGGTGTGACACGAGC--- 1054
Qy 38 AsnAspSerSerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGly 57
Db 1055 -----CATCCGACCTGAAGGGCGGGTGACCGCTCCAGAACTTCACC 1099
Qy 58 ArgThrAsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeu 77
Db 1100 GCCGGCGCGCGCGCGGACAGAGTGGGCCACGGCACCCACGCTCGCTCGATCGCGCG 1159
Qy 78 GlyAsnAlaThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuVal 92
Db 1160 GGCACGGCGCGCGCGAGTCAAGGGCAAGTACAAAGGCGGTGCGACCCGCGCGCGATCCTC 1219
Qy 93 PheGlnSerIleMet---AspSerGly---GlyGlyLeuGlyGlyLeuProAlaAsnLeu 110
Db 1220 AACGCAAGGCTCTCGACGACTCCGGTTTCGGCGACGACTCCGGCATCTCCGCGGCGATG 1279
Qy 111 GlnThrLeuPheSerGlnAlaTyrSerAlaGlyAlaArgIleHisThrAsnSerTrpGly 130
Db 1111 -----

```

RESULT 15  
AAT61455  
ID AAT61455 standard; DNA; 2809 BP.

```

Db 1280 GAGTGGGCGCGCGCGCAG-----GGCGCGGAGCGTCTGTCACATGAGCCTGGGC 1327
Qy 131 Ala-----ProValAsnGlyAlaTyrThrThrAspSerArgAsn 143
Db 1328 GGCATGGACACACCGGAGACCGACCGCTGGAGCGGCG----- 1366
Qy 144 ValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAlaGlyAsnGluGly 163
Db 1367 GTCGACAGCTGTCCGCGAGAGGGGTCTGTTCGCCATCGCGCGCGGCAACGAGGCG 1426
Qy 164 ProGlySerGlyThrIleSerAlaProGlyThrLalysAsnAlaIleThrValGlyAla 183
Db 1427 CCGGAGTGC-----ATCGGTTCGCGCGGACGCGGCGCGCTCACCCTCGCGGCGC 1480
Qy 184 ThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsnHisValAlaGln 203
Db 1481 -----GTCGACGACAGGACAGCTCGCGGAC 1507
Qy 204 PheSerSerArgGlyPro---ThrArgAspGlyArgIleLysProAspValMetAlaPro 222
Db 1508 TTCTCTCCACCGCGCGCGCGCTCGCGACGCGCCATCAAGCGGAGCGTCAACGCTCCC 1567
Qy 223 GlyThrTyrIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrpAlaAsn 242
Db 1568 GCGTGGACATCACGCGCGCTCGCGGAGGCGCAACGACATCGCGCGGAGGTGCGTAG 1627
Qy 243 HisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIleValAlaGly 262
Db 1628 GGACGCGCGCGCTACATGACCATCTCCGCGACGTCGATGGCGACCCCGCAGCTCGCGGC 1687
Qy 263 AsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyValThrProLysProSer 282
Db 1688 GCGCGCGCGCTCTCGAAGCAGCAG-----CACCCCGAC 1720
Qy 283 LeuLeuLysAlaAlaLeu-----IleAlaGlyAlaAlaAspValGlyLeuGlyPhe 299
Db 1721 TGGACCTCGCGCGAAGTGAAGCGCGCTCACCGGCTCCACCAAGGGCGGCAAGTACACC 1780
Qy 300 ProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerSerLeuAsnValAlaPhe 319
Db 1781 CCG---TTGAGCAGGGTTCGCGCGGATCCAGCGCGCAAGGGCTCCAGCAGACCGGTG 1837
Qy 320 ValAsnGluThrSerProLeuSerThrSer----- 329
Db 1838 ATC-----GCCGACCGCGTCTCGGTGAGCTTCGGCGTCCAGCAGTGGCGGACCGAC 1891
Qy 330 -----GlnLysAlaThrTyrSerPheThrAlaGlnAlaGlyLysProLeu 344
Db 1892 GACGAGCGCGTCAACCAAGCAGCTGACCTACCGCAACCTCGGCGCACCCAGGACGTCAGCTG 1951
Qy 345 LysIleSerLeuValTrpSerAsp-----AlaPro----- 354
Db 1952 AAGCTGACGTGACCGCGCGCACCCCAAGGGCAAGGGCGGCGCGCGGGCTTCTTCACG 2011
Qy 355 ---GlySerThrThrAlaSerLeuThrLeuValAsnAspLeuAspLeuValIleThrAla 373
Db 2012 CTGGGCGCGCACCG-----GTGACCGCTC 2035
Qy 374 ProAsnGlyThrLysTyrValGlyAsnAspPheThrAlaProTyrAspAsnAsnTrpAsp 393
Db 2036 CCGCGCGCGCGGCGAGC---GCCCTCCGTCGACATGACCGCGC-----GACACCGCGCTCGGC 2086
Qy 394 GlyArg---AsnAsnValGluAsnValPheIleAsnAlaProGlnSerGlyThrThr 412
Db 2087 GGCACGCTGAGCGCGCGTACTCGCGGTACGTGGTTCGCC---ACGGCGCGCGGCGAGACG 2143
Qy 413 Val-----GluValGlnAlaTyrAsnValProVal 422
Db 2144 GTCCGACGCGCGCGCGGTGACGCGGAGGTGAGTCTGACGACGTGACCGCTC 2197

```



XX AAT61455;  
 AC 17-OCT-2003 (revised)  
 DT 06-OCT-1997 (first entry)  
 XX DhpA-mel chimeric gene.  
 DE  
 XX asymmetric hydrolase; dhpA; 4-substituted-1,4-dihydropyridine;  
 KW derivative; Streptomyces viridosporus; ester; chiral; synthesis;  
 KW cardiovascular; treatment; hypertension; ischaemic heart disease; ds.  
 XX  
 OS Streptomyces viridosporus.  
 OS Streptomyces antibioticus.  
 OS Chimeric.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 338..2809  
 FT /\*tag= a  
 FT mat\_protein 338..2539  
 FT /\*tag= b  
 FT /product= "dhpA protein product"  
 FT /note= "from S. viridosporus"  
 FT mat\_protein 2540..2809  
 FT /\*tag= c  
 FT /product= "Melanin"  
 FT /note= "from S. antibioticus"  
 XX  
 XX WO9705243-A1.  
 XX 13-FEB-1997.  
 XX 30-JUL-1996; 96WO-JP002147.  
 XX 31-JUL-1995; 95JP-00212975.  
 XX 29-FEB-1996; 96JP-00067478.  
 XX (SAOC ) MERCIAN CORP.  
 XX  
 XX Arisawa A, Matsufuji M, Tsuruta T, Dobashi K, Nakashima T;  
 XX Tshiki K, Yoshioka T;  
 XX WPI; 1997-145682/13.  
 XX P-PSDB; AAW13668.  
 XX  
 XX Asymmetric hydrolase gene derived from Streptomyces viridosporus - acts  
 XX on 4-substituted-1,4-dihydro-pyridine derivatives to produce chiral  
 XX derivatives useful for synthesis of cardiovascular drugs.  
 XX  
 XX Claim 5; Page 37-43; 78pp; Japanese.  
 XX  
 XX This sequence is a fusion gene encoding Streptomyces viridosporus dhpA  
 XX gene product (an asymmetric hydrolase) which acts on 4-substituted-1,4-  
 XX dihydropyridine derivatives, and melanin from S. antibioticus. The dhpA  
 XX enzyme allows the efficient conversion of 4-substituted-1,4-  
 XX dihydropyridine esters to chiral partially hydrolysed derivatives, for  
 XX use in the synthesis of cardiovascular drugs suitable for the treatment  
 XX of e.g. hypertension and ischaemic heart disease. (Updated on 17-OCT-2003  
 XX to standardise OS field)  
 XX  
 XX Sequence 2809 BP; 473 A; 1041 C; 984 G; 311 T; 0 U; 0 Other;  
 SQ  
 Alignment Scores:  
 Pred. No.: 1,89e-16 Length: 2809  
 Score: 327.50 Matches: 136  
 Percent Similarity: 40.83% Conservative: 51  
 Best Local Similarity: 29.69% Mismatches: 160  
 Query Match: 14.54% Indels: 111  
 DB: 2 Gaps: 23  
 US-09-985-689A-7 (1-433) x AAT61455 (1-2809)  
 QY 18 GlyLeuTyrGlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArg 37

Db 998 GGCTACGACGGGAGGTGAGATCGCGTCTGGACACCGGTGTCGACAGG--- 1054  
 QY 38 AsnAspSerSerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGly 57  
 Db 1055 -----CATCCGACCTGAAGGGCGGGTGCACCGCGCTCAAGAACTTCACC 1099  
 QY 58 ArgThrAsnAsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeu 77  
 Db 1100 GCGCGCGCGCGCGGCGGACCAAGTGGGCCACCGCACCGCTCGATCGCGCGG 1159  
 QY 78 GlyAsnAlaThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuVal 92  
 Db 1160 GGCACGCGCGCGCGCTCCAGGGGCAAGTACAGGGCGTTCGACCGCGCGCGGATCCTC 1219  
 QY 93 PheGlnSerIleMet---AspSerGly---GlyGlyLeuGlyGlyLeuProAlaAsnLeu 110  
 Db 1220 AACGGCAAGGTCTCTCGACGACTCCGGTTTCGGGACGACTCCGGCATCTCGCGCGCATG 1279  
 QY 111 GlnThrLeuPheSerGlnAlaTyrSerAlaGlyAlaArgIleHisThrAsnSerTrpGly 130  
 Db 1280 GAGTGGCGCGCGCGCAG-----GGCGCGCAGCTGCTCAACATGAGCTCGGC 1327  
 QY 131 Ala-----ProValAsnGlyAlaTyrThrThrAspSerArgAsn 143  
 Db 1328 GGCATGGACACACCGGACCGACCGCTGGAGCGCGG----- 1366  
 QY 144 ValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaGlyAsnGluGly 163  
 Db 1367 GTCGACAAAGCTGTCGCGGAGAGGGGTCTCTGTCGTCATCGCGCGCGGCAAGAGGCG 1426  
 QY 164 ProGlySerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrValGlyAla 183  
 Db 1427 CCGGAGTGG-----ATCGGTTCGCGCGGACGCGGCGCGCTCAGCGGAGCTCAGCGTCC 1480  
 QY 184 ThrGluAsnLeuLeuArgProSerPheGlySerTyrAlaAspAsnIleAsnHisValAlaGln 203  
 Db 1481 -----GTCGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 1507  
 QY 204 PheSerSerArgGlyPro---ThrArgAspGlyArgIleLysProAspValMetAlaPro 222  
 Db 1508 TTCTCTCTCACCGCGCGCGCGCTCGCGGACGCGGCGCATCAAGCGGAGCTCAGCGTCC 1567  
 QY 223 GlyThrTyrIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrpAlaAsn 242  
 Db 1569 GCGGTGGACATCACGCGCGCTCGCGGAGGGCAACGACATCGCGCAGGAGGTGCGTGAG 1627  
 QY 243 HisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIleValAlaGly 262  
 Db 1628 GACCGCGCGCGCTACATGACCATCTCCGCGCAGCTCGATGGCGACCGCGCGCTCGCGGC 1687  
 QY 263 AsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyValThrProLysProSer 282  
 Db 1688 GCGCGCGCGCTCTCGACGACGAG-----CACCCCGAC 1720  
 QY 283 LeuLeuLysAlaAlaLeu-----IleAlaGlyAlaAlaAspValGlyLeuGlyPhe 299  
 Db 1721 TGGACCTCGCGCGAACTGAAGGGCGCGCTCACCGGCTCCACCAAGGGGCGCAATACACC 1780  
 QY 300 ProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaPhe 319  
 Db 1781 CGG---TTCGACGAGGTTCGGCGCGATCCAGCGCGACAAAGCGCTCCAGCAGACCGTG 1837  
 QY 320 ValAsnGluThrSerProLeuSerThrSer----- 329  
 Db 1838 ATC-----GCGGACCGCGTCTCGGTGAGCTTCGCGCTCCAGCAGTGGCGCGCACCGGAC 1891  
 QY 330 -----GlnLysAlaThrThrSerPheThrAlaGlnAlaGlyLysProLeu 344  
 Db 1892 GACGAGCGGTCCACAGCAGCTGACCTACCGCAACCTCGGCACCGCAGGAGCTCAGCGTG 1951  
 QY 345 LysIleSerLeuValTrpSerAsp-----AlaPro----- 354

1552	Db	AACTGACGTGACCGCCACCGACCGACCGGCAAGGGCAAGCGGCCCGCGGGGCTTCTTCACG	2011
355	Qy	---GlySerThrThrAlaSerLeuThrLeuValAsnAspLeuAspLeuValIleThrAla	373
2012	Db	CTGGCGGCCACCCAGC-----GTGACCGTC	2035
374	Qy	ProAsnGlyThrIleTyrValGlyAsnAspPheThrAlaProTyrAspIleAsnIleTrpAsp	393
2036	Db	CCGCGCGGGCGGACG---GCCTCGTCGCATACCGCC-----GACACCGCGGCTCGGC	2086
394	Qy	GlyArg---AsnAsnValGluAsnValPheIleAsnAlaProGlnSerGlyThrTyrThr	412
2087	Db	GGCACGGTGACGGCGCGGTACTCGGCGTACGTGGTCGCC---ACGGGGCGGGCGACACG	2143
413	Qy	Val-----GluValGlnAlaTyrAsnValProVal	422
2144	Db	GTCCGACGCGCGCGCGGGTGCACGGCGAGGTGCAGTCTGACACGTGACCGTC	2197

Search completed: April 4, 2004, 01:13:19  
Job time : 363.908 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: April 3, 2004, 16:47:19 ; Search time 3544.27 Seconds  
(without alignments)  
5295.174 Million cell updates/sec

Title: US-09-985-689A-7

Perfect score: 2252

Sequence: 1 NDVARGIVKADVAQNFGLY.....EVQAYNVVPSPQTFSLAIVH 433

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame+\_p2n.model -DEV=xlh  
-O=/cgn2\_1/USPTO\_epool/US09985689/runat\_31032004\_161806\_4155/app.query.fasta\_1.3498  
-DB=GenEmbl -QFWT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFWT=plo -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09985689@cgn\_1\_1\_1925\_@runat\_31032004\_161806\_4155 -NCPU=6 -ICPU=3  
-NO\_MAP -LARGESQUERY -NEG\_SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

1: gb.ba.\*  
2: gb.htg.\*  
3: gb.in.\*  
4: gb.om.\*  
5: gb.ov.\*  
6: gb.pat.\*  
7: gb.ph.\*  
8: gb.pl.\*  
9: gb.pr.\*  
10: gb.ro.\*  
11: gb.sts.\*  
12: gb.sy.\*  
13: gb.un.\*  
14: gb.vi.\*  
15: em.ba.\*  
16: em.fun.\*  
17: em.hum.\*  
18: em.in.\*  
19: em.ml.\*  
20: em.om.\*  
21: em.or.\*  
22: em.ov.\*  
23: em.pat.\*  
24: em.ph.\*  
25: em.pl.\*  
26: em.ro.\*  
27: em.sts.\*  
28: em.un.\*

29: em.vi.\*  
30: em.htg\_hum.\*  
31: em.htg\_inv.\*  
32: em.htg\_other.\*  
33: em.htg\_mus.\*  
34: em.htg\_pla.\*  
35: em.htg\_rod.\*  
36: em.htg\_mam.\*  
37: em.htg\_vrt.\*  
38: em.ey.\*  
39: em.htgo\_hum.\*  
40: em.htgo\_mus.\*  
41: em.htgo\_other.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	2252	100.0	3003	6	AR089954 Sequence
2	2252	100.0	3003	6	BD062155 Nucleic a
3	2130.5	94.6	1923	1	AB084155 Bacillus
4	2125.5	94.4	1305	6	AX839476 Sequence
5	2125.5	94.4	1923	1	AB051423 Bacillus
6	2125.5	94.4	1923	6	AR368117 Sequence
7	2120.5	94.2	1923	6	AR368118 Sequence
8	2116.5	94.0	1302	1	AB046406 Bacillus
9	2110.5	93.7	1920	1	AB046403 Bacillus
10	2110.5	93.7	1920	6	AR368116 Sequence
11	2044	90.8	1299	1	AB046405 Bacillus
12	2040	90.6	1299	1	AB046402 Bacillus
13	2033	90.3	1299	1	AB046404 Bacillus
14	2032	90.2	2218	6	E03808
15	626.5	27.8	60005	1	AF268611 Uncultured
16	568.5	25.2	301205	1	AE017218 Geobacter
17	501	22.2	5772	3	AF466309 Dictyoste
18	499	22.2	5890	3	DDU60086 Dictyostell
19	494.5	22.0	6115	3	DDU20432 Dictyostell
20	445.5	19.8	5109	3	AF263455 Dictyoste
21	432.5	19.2	1977	6	AR201152 Sequence
22	423	18.8	126928	2	AC096673 Trypanoso
23	404	17.9	14376	1	AE013049 Thermoana
24	396	17.6	1235	6	AR201146 Sequence
25	396	17.6	1235	6	AR201155 Sequence
26	396	17.6	12452	1	AE010265 Pyrococcu
27	377	16.7	299850	1	AP004601 Oceanobac
28	374.5	16.6	303450	1	SC0939130 Streptomy
29	355	15.8	300425	1	AP005044 Streptomy
30	341.5	15.2	3348	1	D83672 Streptomyce
31	339.5	15.1	1239	1	AF305633 Thermoana
32	339.5	15.1	1239	1	AY028704 Thermoana
33	333.5	15.1	12295	1	AE013026 Thermoana
34	334.5	14.9	292550	1	AP001513 Bacillus
35	332.5	14.8	2539	6	AR202322 Sequence
36	332	14.7	302300	1	AP005034 Streptomy
37	329.5	14.6	291000	1	SC0939105 Streptomy
38	327.5	14.5	2809	6	AR202321 Sequence
39	327.5	14.5	3900	1	AB007809 Streptomy
40	326.5	14.5	300800	1	SC0939112 Streptomy
41	325	14.4	1329	6	AX433519 Sequence
42	323.5	14.4	207829	1	BSU80010 Bacillus su
43	323	14.3	135638	1	AF484556 Streptomy
44	322	14.3	6854	1	BSU99230 Bacillus sp
45	320	14.2	2731	1	AF015225 Thermococ

# ALIGNMENTS

RESULT 1

```
AR069954
LOCUS       AR069954
DEFINITION   Sequence 41 from patent US 5891701.
ACCESSION   AR069954
VERSION     AR069954.1 GI:7220842
KEYWORDS    Unknown.
SOURCE      Unknown.
ORGANISM    Unclassified.
REFERENCE   1 (bases 1 to 3003)
AUTHORS     Sloma,A. and Christianson,L.
TITLE       Nucleic acids encoding a polypeptide having protease activity
JOURNAL     Patent: US 5891701-A 41 06-APR-1999;
            Location/Qualifiers
FEATURES             1..3003
                     /organism="unknown"
                     /mol_type="unassigned DNA"

ORIGIN
Alignment Scores:
Pred. No.:      2,23e-133      Length:      3003
Score:          2252.00        Matches:      433
Percent Similarity: 100.00%    Conservative: 0
Best Local Similarity: 100.00% Mismatches:      0
Query Match:     100.00%      Indels:         0
DB:              6            Gaps:           0

US-09-985-689A-7 (1-433) x AR069954 (1-3003)
QY      1 AsnAspValAlaAArgGlyVileVallyeAlaAaspValAlaGlnAenAenPheGlyLeuTyr 20
DB      1470 ANTACGTCGGCCGTGGCATTTGTGAAGCAGACGTCGCACAAAATAACTTTGGCTTATAT 1529
QY      21 GlyGlnGlyGlnIleValAlaValAlaAaspThrGlyLeuAaspThrGlyArgAenAaspSer 40
DB      1530 GGACAAGACAGAGATTGTAGCAGTTGCTGATATCGGCTTGATACAGAAAGAAATGACGT 1569
QY      41 SerMetHisGluAlaPheAArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
DB      1590 TCGATGCAATGAGCATTCGCGGTAGATTACCGCATATATGCTGCTGGCAGAGCAAGAT 1649
QY      61 AsnAlaAenAaspProAasnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnAla 80
DB      1650 AACGCCAATGATCCAAATGGACATGGAAACCCCATGTTGCTGGATCTGTTAGGAAATGCT 1709
QY      81 ThrAenLysGlyMetAlaProGlnAlaAenLeuValPheGlnSerIleMetAaspSerGly 100
DB      1710 ACAATAAAGGATGGCACCAGCCCAATCTAGCTTTCAATCTATTATGATAGTGGT 1769
QY      101 GlyGlyLeuGlyGlyLeuProAlaAenLeuGlnThrLeuPheSerGlnAlaTyrSerAla 120
DB      1770 GGAGGGCTGGAGGACTACCTGCTAATCTACAAACATTTATTCAGTCAAGCATATAGTGCT 1829
QY      121 GlyAlaAArgIleHisThrAasnSerTrpGlyAlaProValAenGlyAlaTyrThrThrAsp 140
DB      1830 GGAGCGAAGATTCATACCAATTCATGGGGGCTCCAGTAAACGGTGCTTATACAGACAC 1889
QY      141 SerA-GAsnValAaspAaspTyrValArgLysAenAaspMetThrIleLeuPheAlaAlaGly 160
DB      1890 TCTCGAATGTTGATGATTATGTGAGAAAATAATGATATGACGATCTTTTTCGGCGCGGA 1949
QY      161 AsnGluGlyProGlySerGlyThrIleSerAlaProGlyThrAlaLysAenAlaIleThr 180
DB      1950 AATGAGGGACCAAGTAGCGGTACATTCAGTCAGCAGGACAGCAAGAAAATGCGATTACA 2009
QY      181 ValGlyAlaThrGluAenLeuAArgProSerPheGlySerTyrAlaAaspAasnIleAasnHis 200
DB      2010 GTTGGGGCAACCGAAACCTAGTCCAGCTTCGATCTTATCGGATTAATATTAAACCAT 2069
QY      201 ValAlaGlnPheSerArgGlyProThrArgAaspGlyArgIleLysProAaspValMet 220
DB      2070 GTTGTCTCAATCTCTTCACGAGTCTCTACTAGAGATGGACGATTATTAAGCCGACGTCATG 2129
```

```
QY      221 AlaProGlyThrTyrIleLeuSerAlaAArgSerSerLeuAlaProAaspSerSerPheTyr 240
DB      2130 GCACACAGTACGTATATCTCTCTGTAGATCATCATTAGCTCCAGATTCCTATTTCTGG 2189
QY      241 AlaAenHisAaspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIleVal 260
DB      2190 GCARACCATGATAGTAATATGCCTACATGGTGGTACTTCTATGGCTACTCCAAATTGTA 2249
QY      261 AlaGlyAenValAlaGlnLeuAArgGluHisPheVallyeAenArgGlyValThrProLys 280
DB      2250 GCAGGTAATGTGTGCACAATTAAGGGAGCATTTTGTGAAATAATAGAGGGTAACCTCTTAAG 2309
QY      281 ProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAaspValGlyLeuGlyPhePro 300
DB      2310 CTTTCCCTTTTAAAGCTGCTTTAATTGCAGTCTCTGGGATGTTGGACTTGGCTTTCCA 2369
QY      301 AenGlyAenGlnGlyTrpGlyArgValThrLeuAaspLysSerLeuAenValAlaPheVal 320
DB      2370 AATGGTAACCAAGGATGGGAAGAGTAACGTTAGATAAATCCCTAAATGTCGCAATTTGTG 2429
QY      321 AenGluThrSerProLeuSerThrSerGlnLysAlaThrTyrSerPheThrAlaGlnAla 340
DB      2430 AATGAACAGACCCCTTTATCAACAGTCACAAAGCAACATATTCTGTACGGCTCAAGCT 2489
QY      341 GlyLysProLeuLysIleSerLeuValTrpSerAaspAlaProGlySerThrThrAlaSer 360
DB      2490 GGTAAACCTTTAAATAATATCACTTTGTTGGTCAGATGCACAGGTAGCACACGCGATCA 2549
QY      361 LeuThrLeuValAasnAaspLeuAaspLeuValIleThrAlaProAasnGlyThrLysTyrVal 380
DB      2550 CTAACTTTAGTGAATGATTTAGATTAGATTATCTACCTGACCAATGGAACCTAAATACGTC 2609
QY      381 GlyAenAaspPheThrAlaProTyrAaspAenAenTrpAaspGlyArgAasnValGluAen 400
DB      2610 GGAAATGACTTTTACAGCACCGGTATGATAACAATTTGGATGGCAGAAACAACGTTGGAAT 2669
QY      401 ValPheIleAasnAlaProGlnSerGlyThrTyrThrValGluValGlnAlaTyrAasnVal 420
DB      2670 GTGTTTATCAATGCTCTCTCAAGCGGAACGATATACGTCGAAGTCAGAGCTTACATGTA 2729
QY      421 ProValSerProGlnThrPheSerLeuAlaIleValHis 433
DB      2730 CCAGTAAGTCCGCAACCTTTTCTTTAGCGATTGTACAT 2768

RESULT 2
BD062155      3003 bp      DNA      linear      PAT 27-AUG-2002
LOCUS         Nucleic acids encoding a polypeptide having protease activity.
DEFINITION
ACCESSION     BD062155
VERSION       BD062155.1 GI:22607760
KEYWORDS      JP 2001514529-A/39.
SOURCE        Homo sapiens (human)
ORGANISM      Homo sapiens
REFERENCE     1 (bases 1 to 3003)
AUTHORS       Sloma,A. and Christianson,L.
TITLE         Nucleic acids encoding a polypeptide having protease activity
JOURNAL       Patent: JP 2001514529-A 39 11-SEP-2001;
            NOVO NORDISK BIOTECH INC
            PN JP 2001514529-A/39
COMMENT       PD 11-SEP-2001
            PF 09-JUN-1998 JP 1999503145
            PR 12-JUN-1997 US 08/873479
            PI ALAN SLOMA,LYNNE CHRISTIANSON
            PC C12N15/57,C12N15/75,C12N9/54,C12K14/00
            CC Strandedness: Single;
            CC Topology: Linear;
            PH Key Location/Qualifiers.
FEATURES             1..3003
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
```

/db\_xref="taxon:9606"

## ORIGIN

## Alignment Scores:

Pred. No.: 2,238-133 Length: 3003  
 Score: 2252.00 Matches: 433  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 6 Gaps: 0

US-09-985-689a-7 (1-433) x BD062155 (1-3003)

QY 1 AsnAspValAlaArgGlyLeuValLysAlaAspValAlaGlnAsnAspPheGlyLeuTyr 20  
 DB 1470 AATGACGTCGCGCGTGGCAATCTGTAAGAGCAGACGTCGCACAAAATAAATCGCTTATAT 1529  
 QY 21 GlyGlnGlyGlnLeuValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40  
 DB 1530 GGAACAGGACAGATTTAGCAGTTCTGATACCTGGCTTGATACAGGAAGAAATGACAGT 1589  
 QY 41 SerMetHisGluAlaPheArgGlyLysLeuThrAlaLeuTyrAlaLeuGlyArgThrAsn 60  
 DB 1590 TCGATGCATGAAGCATTCGCGGTAAAGATTACCGCACTATATGCACTGGGCGAAGCAAT 1649  
 QY 61 AsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnAla 80  
 DB 1650 AACGCCAATGATCCAAATGACATGGAACCCATGTTGCTGATCTGTTAGGAATGCT 1709  
 QY 81 ThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSerGly 100  
 DB 1710 ACAAAATAAAGGATGACCGCAAGCAATCTAGTCTTTCATCTTATATGATGATGCT 1769  
 QY 101 GlyGlyLeuGlyGlyLeuProAlaAsnLeuGlnThrLeuPheSerGlnAlaTyrSerAla 120  
 DB 1770 GAGGAGCTGGGAGGACTACTCTGTAATCTACAAACATTTATGATGACGATATGCT 1829  
 QY 121 GlyAlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyrThrThrAsp 140  
 DB 1830 GAGAGCAGAAATTCATACGAATTCATGCGGGGCTCCAGTAACGGTCCCTATACACAGAC 1889  
 QY 141 SerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAlaGly 160  
 DB 1890 TCTCGAAATGTTGATGATTTATGTGAGAAATAATGATATGACGATCTTTTGGCGCGGA 1949  
 QY 161 AsnGluGlyProGlySerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThr 180  
 DB 1950 AATGAGGAGCAGTAGCGGTCAATCAGTCAGCCAGGACAGCAAAAATGCGATTACA 2009  
 QY 181 ValGlyValaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsnHis 200  
 DB 2010 GTTGGGGCAACCGAAACCTAGTCCAGCTTCGATCTTATGCGGATATATTAACCAT 2069  
 QY 201 ValAlaGlnPheSerSerArgGlyProThrArgAspGlyArgIleLysProAspValMet 220  
 DB 2070 GTTGCTCAATCTCTTCAGAGGTCTCTACTAGATGGACGATTAATGAGCGGACGTCATG 2129  
 QY 221 AlaProGlyThrTyrIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTyr 240  
 DB 2130 GCACCAAGTACGATATATCTCTGCTAGATCATCATTAGCTCCAGATCTCTCATTTCTGG 2189  
 QY 241 AlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIleVal 260  
 DB 2190 GCAACCATGATAGTAATATCCCTACATGGGTGATCTTCTATGCTTACTCCAAATGTA 2249  
 QY 261 AlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyValThrProLys 280  
 DB 2250 GCAGGTAAATGTTGCACAAATTAAGGAGCATTTTGTGAAAATATAGGGGGAACCTTAAG 2309  
 QY 281 ProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspValGlyLeuGlyPhePro 300  
 DB 2310 CCTTCCCTTTAAAGCTCTTTAATTGAGGTGCTGGGATGTTGGACTTTGGCTTTTCCA 2369

QY 301 AsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaPheVal 320  
 DB 2370 AATGTATACCAAGATGGGAAGAGTAACTGTTAGATTAATCCCTAATATGCGGATTTGTG 2429  
 QY 321 AsnGluThrSerProLeuSerThrSerGlnLysAlaThrTyrSerPheThrAlaGlnAla 340  
 DB 2430 AATGAAACGAGCCCTTTATCAACAAGTCAAAAAGCAACATATTCTGTTACGGCTCAAGCT 2489  
 QY 341 GlyLysProLeuLysIleSerLeuValTyrSerAspAlaProGlyLysThrThrAlaSer 360  
 DB 2490 GGTAAACCCCTTAAATATATCATTTGTTGTCAGATGCACACGATGACGACGCGATCA 2549  
 QY 361 LeuThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrLysTyrVal 380  
 DB 2550 CTAATCTTTAGTGAATGATTAGACTTTAGTATATCACTGCACCAATGGAACATAATACGTC 2609  
 QY 381 GlyAsnAspPheThrAlaProTyrAspAsnAspTyrAspGlyArgAsnAsnValGluAsn 400  
 DB 2610 GGAATGACTTTACAGCACCGTATGATTAACAATGGATGCGAGAACACCTGGGAAAT 2669  
 QY 401 ValPheIleAsnAlaProGlnSerGlyThrTyrThrValGluValGlnAlaTyrAsnVal 420  
 DB 2670 GTGTTTATCAATGCTCTCTCAAGCGGAACGTATACAGTCGAACTGACGGCTTACAATGTA 2729  
 QY 421 ProValSerProGlnThrPheSerLeuAlaIleValHis 433  
 DB 2730 CCAAGTATGCTCGCAACCTTTCTTTAGCGATTGTACAT 2768

## RESULT 3

AB084155

LOCUS

DEFINITION

AB084155

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

Source

CDS

FEATURES

Source

CDS

FEATURES

Source

CDS

FEATURES

Source

CDS

FEATURES

Source

CDS

FEATURES

Source

CDS

FEATURES

Source

CDS

FEATURES

Source

CDS

FEATURES

Source

CDS

FEATURES

Source

CDS

FEATURES

Source

CDS

FEATURES

Source

CDS

FEATURES

Source

CDS

FEATURES

Source

CDS

FEATURES

Source

CDS

FEATURES

Source

CDS

FEATURES

Source

CDS

FEATURES

Source

CDS

FEATURES

Source

CDS

FEATURES

Source

CDS

FEATURES

Source

CDS

FEATURES

Source

CDS

FEATURES

Source

CDS

FEATURES

Source

CDS

FEATURES

Source

CDS

FEATURES

Source

CDS

FEATURES

Source

CDS

FEATURES

Source

CDS

FEATURES

Source

CDS

FEATURES

Source

CDS

FEATURES

Source

CDS

FEATURES

Source

CDS

FEATURES

Source

CDS

FEATURES

Source

CDS

FEATURES

Source

CDS

FEATURES

Source

CDS

FEATURES

Source

CDS

FEATURES

Source

CDS

FEATURES

Source

CDS

FEATURES

Source

CDS

FEATURES

Source

CDS

FEATURES

Source

CDS

FEATURES

Source

CDS

FEATURES

Source

CDS

FEATURES

Source

CDS

FEATURES

Source

CDS

FEATURES

Source

CDS

FEATURES

Source

CDS

FEATURES

Source

CDS

FEATURES

Source

CDS

FEATURES

Source

CDS

FEATURES

Source

CDS

FEATURES

Source

CDS

FEATURES

Source

CDS

FEATURES

Source

CDS

FEATURES

Source

CDS

FEATURES

Source

CDS

FEATURES

Source

CDS

FEATURES

Source

CDS

FEATURES

Source

CDS

FEATURES

Source

CDS

FEATURES

Source

CDS

FEATURES

Source

CDS

FEATURES

Source

CDS

FEATURES

Source

CDS

FEATURES

Source

CDS

FEATURES

Source

CDS

FEATURES

Source

CDS

FEATURES

Source

CDS

FEATURES

Source

CDS

FEATURES

Source

CDS

FEATURES

Source

CDS

FEATURES

Source

CDS

FEATURES

Source

CDS

FEATURES

Source

CDS

FEATURES

Source

CDS

FEATURES

Source

CDS

FEATURES

Source

CDS

FEATURES

Source

CDS

FEATURES

Source

CDS

FEATURES

Source

CDS

FEATURES

Source

CDS

FEATURES

Source

CDS

FEATURES

Source

CDS

FEATURES

Source

CDS

FEATURES

Source

CDS

FEATURES

Source

CDS

FEATURES

Source

CDS

FEATURES

Source

CDS

FEATURES

Source

CDS

FEATURES

Source

CDS

FEATURES

Source

CDS

FEATURES

Source

CDS

FEATURES

Source

CDS

FEATURES

Source

ORIGIN

Alignment Scores:  
Pred. No.: 6.3e-126 Length: 1923  
Score: 2130.50 Matches: 407  
Percent Similarity: 97.93% Conservativeness: 18  
Best Local Similarity: 93.78% Mismatches: 8  
Query Match: 94.60% Indels: 1  
DB: 1 Gaps: 1

US-09-985-689A-7 (1-433) x AB084155 (1-1923)

Qy 1 AsnAspValAlaArgGlyIleValIleValAlaAspValAlaGlnAsnAsnPheGlyLeuTyr 20  
Db 619 AATGATGTTGACGTCGGAATTTGTAAGCGGATGTCGCGAGACGACGACGAGGTTGAT 678  
Qy 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40  
Db 679 GGACAGGACAGATCGTAGCGGTTGCCGATACAGGCTTGATACAGGTCGCAATGACAGT 738  
Qy 41 SerVecHisGluAlaPheArgGlyIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60  
Db 739 TCGATGATGAAGCTTCCCGGGGAAATTAATGATGATGATGATGATGATGATGATGAT 798  
Qy 61 AsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn--- 79  
Db 799 AATGCCAATGATACGAATGGTACGATGATGATGATGATGATGATGATGATGATGATG 858  
Qy 80 AlaThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 99  
Db 859 TCCACTAATAAGGAATGGCCCTCAGCGAATCTAGCTTCCCAATCTATCATGATGATG 918  
Qy 100 GlyGlyGlyLeuGlyGlyLeuProAlaAsnLeuGlnThrLeuPheSerGlnAlaTyrSer 119  
Db 919 GGTGGGGACTTGGAGGACTACCTTCGATCTGCAACCTTATTTCAGCAGCATACAGT 978  
Qy 120 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyrThr 139  
Db 979 GCTGTGTCAGAAATTCATACAACTCTGGGAGCAGAGTGAATGGGCTTTACACAACA 1038  
Qy 140 AspSerArgAsnValAlaAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 159  
Db 1039 GATTCCAGAAATGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1098  
Qy 160 GlyAsnGluGlyProGlySerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 179  
Db 1099 GGGATGAAGGACCGAAGCGGACCATCAGTCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1158  
Qy 180 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 199  
Db 1159 ACAGTCGGAGCTACGGAAACCTTCCGCCCAAGCTTCCGGTCTTATGCGGACATATCAC 1218  
Qy 200 HisValAlaGlnPheSerArgGlyProThrArgAspGlyArgIleLysProAspVal 219  
Db 1219 CATGTGGCAGATCTCTTCACGTGACCGACGACAAAGGATGAGCGGATCAACCGGATG 1278  
Qy 220 MetAlaProGlyThrTyrIleLeuSerAlaArgSerSerLeuAlaProAspSerPhe 239  
Db 1279 ATGGCACCAGGACGTTCACTATCATCAGCAAGATCTTCTTTCGACCGGATTCCTCTTC 1338  
Qy 240 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProIle 259  
Db 1339 TGGCGGAACCATGACAGTAATATGATATGATGATGATGATGATGATGATGATGATG 1398  
Qy 260 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyValThrPro 279  
Db 1399 GTTGTGGAAACCTGGCAGCTTCGTGAGCATTTTGTGAAGAACAGAGGATCACACCA 1458  
Qy 280 LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspValGlyLeuGlyPhe 299  
Db 1459 AAGCTTCTCTATTAAAGCGGACGATGATGCGCGGTGCAGCAGCATCGCCCTTGGCTAC 1518  
Qy 300 ProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaPhe 319

Db 1519 CCGAACGGTAACCAAGGATGGGACGAGTGACATTTGATAAATCCCTAAACGTTGCTAT 1578  
Qy 320 ValAsnGluThrSerProLeuSerThrSerGlnLysAlaThrTyrSerPheThrAlaGln 339  
Db 1579 GTGAACGAGTCCAGTTCTTATCCACGACGCAAAAGGACGAGTACTGTTTACTGCTACT 1638  
Qy 340 AlaGlyLysProLeuLysIleSerLeuValTrpSerAspAlaProGlySerThrAla 359  
Db 1639 CCGCGCAAGCCTTTGAAGATCTCCCTGATGTTGATGCTGCTGCGGAGCACAACTGCT 1698  
Qy 360 SerLeuThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrLysTyr 379  
Db 1699 TCCGTAACGCTGTCAATGACCTGACCTGCTCATTACCGCTCCAAATGGCACAATAT 1758  
Qy 380 ValGlyAsnAspPheThrAlaProTyrAspAsnAsnTrpAspGlyArgAsnValGlu 399  
Db 1759 GTTGGAAATGACTTACTTCGCCATACAAATAAATGCGGATGCGCGCAATAAGTAGAA 1818  
Qy 400 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrValGluValGlnAlaTyrAsn 419  
Db 1819 AATGATATTATTAAATGCGCCACAAAGCGGACGATACAAATGAGGTACAGCTTATAAC 1878  
Qy 420 ValProValSerProGlnThrPheSerLeuAlaIleValHis 433  
Db 1879 GTACCGGTTGGACACAGACCTTCCTGTTGCAATTGTGAAC 1920  
RESULT 4  
AX839476 1305 bp DNA linear PAT 15-DEC-2003  
LOCUS  
DEFINITION  
Sequence 2 from Patent EP1347044.  
ACCESSION  
AX839476  
VERSION  
AX839476.1 GI:39922766  
KEYWORDS  
Bacillus sp. KSM-KP43  
SOURCE  
Bacillus sp. KSM-KP43  
ORGANISM  
Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
REFERENCE  
1 Okuda, M.K., Sato, T.K., Saito, K.K., Sumitomo, N.K., Izawa, Y.K.,  
AUTHORS  
Saeiki, K.K., Kobayashi, T.K. and Nomura, M.K.  
TITLE  
Alkaline protease  
Patent: EP 1347044-A 2 24-SEP-2003;  
JOURNAL  
Kao Corporation (JP)  
FEATURES  
Location/Qualifiers  
source  
1. .1305  
/organism="Bacillus sp. KSM-KP43"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:109322"  
CDS  
1. .1305  
/notes="unnamed protein product"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CA885146.1"  
/db\_xref="GI:39922767"  
/translation="NDVARGIVKADVAQSSYGLYGQIVAVADTGLDTRNDSSMHE  
AFRKITALYALGRTNANDTGHGTHVAGSVLNGSTKGNAPQALVFSQIMDSGG  
GLGLSPNQLTFSQAYSGARHTNSGAAVNGATITDSNNVDDTVRNDNTLLPA  
GNSGPNQGTISAPGTAKNATIGVATENRSPFGSYADTNHNVAQFSRGTQDKRKP  
DVNAPGTFILSARSLSAPDSFWANHDSKYAMGTSNATPIVAGVQLRHFVKR  
KITPKSLKALIAAGADIGLYPNGNOCWGRVTLDKSLNVAIVNSSLSTSQAT  
YSTATAGPKLTLVSWSDAPASTTASVTLVNDLVIITAPNGTQVGVNDFTSPYNDN  
WDGRNNVENVFAPSGTITIEVQYVNPVPGQTFSLAIVN"

ORIGIN

Alignment Scores:  
Pred. No.: 8.11e-126 Length: 1305  
Score: 2125.50 Matches: 406  
Percent Similarity: 97.93% Conservativeness: 19  
Best Local Similarity: 93.55% Mismatches: 8  
Query Match: 94.38% Indels: 1  
DB: 6 Gaps: 1

US-09-985-689A-7 (1-433) x AX839476 (1-1305)

QY 1 AsnAspValAlaArgGlyLeuVallyAlaAspValAlaGlnAsnAsnPhGlyLeuTyr 20  
 Db 1 AATGATGTTGGCGGTGAATTTCTCAAGCGGATGTGGCTCAGACGAGCTACGGGTTGTAT 60  
 QY 21 GlyGlnGlyGlnLeuValAlaAlaAspThrGlyLeuAspThrGlyValAsnAspSer 40  
 Db 61 GGACAGGACAGATCGTAGCGGTTGCCGATACAGGGCTTGATACAGGTCGCAATGACAGT 120  
 QY 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60  
 Db 121 TCGATGATGAAGCCCTCCCGCGGAAATTAATGCAATATATATGCAATTTGGGACGACCAAT 180  
 QY 61 AsnAlaAsnAspProAsnGlnHisGlyThrHisValAlaGlySerValIleuGlyAsn--- 79  
 Db 181 AATGCCAATGATACGATGGTCAATGATGATGATGATGATGATGATGATGATGATGATGAT 240  
 QY 80 AlaThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 99  
 Db 241 TCCACTAATAAAGGAATGGCGCTCAGGCGAATCTAGTCTTCCAATCTATCATGGATAGC 300  
 QY 100 GlyGlyGlyLeuGlyGlyLeuProAlaAsnLeuGlnThrLeuPheSerGlnAlaTyrSer 119  
 Db 301 GTGGGGGACTGGAGGACTACTCTCGAATCTGCAAACTTATTCAGCCAGCATACAGT 360  
 QY 120 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyrThr 139  
 Db 361 GCTGTGTCAGAAATTCATAAACTCCTGGGAGCAGCAGTGAATGGGGCTTACACACA 420  
 QY 140 AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 159  
 Db 421 GATTCAGAAATGTGGATGACTATGTGGCAAAATGATATGAGATCTTTTCGCTGCC 480  
 QY 160 GlyAsnGlyGlyProGlySerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 179  
 Db 481 GGGAAATGAAGGACCGAAGCGGCAACCATCATGTCACCGAGCAGCAGTGAATAATGCAATA 540  
 QY 180 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 199  
 Db 541 ACAGTCGAGACTACGGAAACCTCCGCCNAGCTTTGGTCTTATGCGGACATATCAAC 600  
 QY 200 HisValAlaGlnPheSerSerArgGlyProThrArgAspGlyA-gileLysProAspVal 219  
 Db 601 CATGTGGCACAGTCTCTTTCAGTGGACCGAAGAGATGGACGGATCAACCGGATGTC 660  
 QY 220 MetAlaProGlyThrTyrIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 239  
 Db 661 ATGGCACCGGAAAGCTTCTACTATCAGCAAGATCTTCTTTGACCGGATCTCTCTTC 720  
 QY 240 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyLysThrSerMetAlaThrProile 259  
 Db 721 TGGGGCAACCATGACAGATAAATATGATACATGATGGGTGGAGAGCTCCATGGCTACACCGATC 780  
 QY 260 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyValThrPro 279  
 Db 781 GTTGTGGAAAGTGGCACAGTCTGTGAGCATTTTGTGAAAGAACAGAGGATCACACCA 840  
 QY 280 LysProSerLeuLeuLysAlaAlaLeuLeuAlaGlyAlaAlaAspValGlyLeuGlyPhe 299  
 Db 841 AAGCCCTCTCTATTAAGCGGCACTGATGTCGGGTGGAGTGGATGATCGGCTTGGGTAC 900  
 QY 300 ProAsnGlyAsnGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsnValAlaPhe 319  
 Db 901 CCGAAGCGTAAACAGGATGGGAGGAGTGCATTTGGATTAATCTCTGGAACGTTGCCAT 960  
 QY 320 ValAsnGlnThrSerProLeuSerThrSerGlnLysAlaThrTyrSerPheThrAlaGln 339  
 Db 961 GTGAACGAGTCCAGTCTCTATCCACCGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1020  
 QY 340 AlaGlyLysProLeuLysLysSerLeuValTrpSerAspAlaProGlySerThrThrAla 359  
 Db 1021 GCCGGCAAGCCCTTGAATAATCTCCCTGGTATGCTGATGCTGCTGCTGCTGCTGCTGCTGCT 1080

QY 360 SerLeuThrLeuValAsnAspLeuAspLeuValIleThrAlaProGlnGlyThrLysTyr 379  
 Db 1081 TCCGTAAACGCTTGTCAATGATCTGACCTTGTCTATACCGCTCCAAATGGCACAGTAT 1140  
 QY 380 ValGlyAsnAspPheThrAlaProTyrAspAsnAsnTrpAspGlyArgAsnAsnValGlu 399  
 Db 1141 CTAGGAATGACTTACTTCGCCATACAAATGATAACTGGATGGCCGCAATACGTAGAA 1200  
 QY 400 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrValGluValGlnAlaTyrAsn 419  
 Db 1201 AATGTATTATTATATGACCAAAAGCGGACGCTATACAAATGAGGTACAGGCTTAAAC 1260  
 QY 420 ValProValSerProGlnThrPheSerLeuAlaIleValHis 433  
 Db 1261 GTACCGGTGGACACAGACCTTCTCGTTGGCAATGTGAAT 1302

RESULT 5  
 LOCUS AB0511423 1923 bp DNA linear BCT 10-MAY-2002  
 DEFINITION Bacillus sp. KP43 PROF gene for protease, complete cds.  
 ACCESSION AB0511423  
 VERSION AB0511423.2 GI:205211154  
 KEYWORDS  
 SOURCE Bacillus sp. KSM-KP43  
 ORGANISM Bacillus sp. KSM-KP43  
 Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
 REFERENCE 1 Itoh, S. and Saeki, K.  
 AUTHORS new protease  
 TITLE 2 (bases 1 to 1923)  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 1923)  
 AUTHORS Saeki, K.  
 TITLE Direct Submission  
 JOURNAL Submitted (21-NOV-2000) Katsuhisa Saeki, KAO CORPORATION;  
 2606, AKABANE, ICHIKAIMACHI, HAGA, TOCHIGI 321-3486, Japan  
 (E-mail:387185@kaxanet.kao.co.jp, Tel.:81285687471 (ex.7471),  
 Fax:81285687403)  
 COMMENT On May 9, 2002 this sequence version replaced gi:14164344.  
 FEATURES  
 Location/Qualifiers  
 1..1923  
 /organism="Bacillus sp. KSM-KP43"  
 /mol\_type="genomic DNA"  
 /strain="KP43"  
 /db\_xref="taxon:109322"  
 1..1923  
 /gene="PROF"  
 1..1923  
 /gene="PROF"  
 /codon\_start=1  
 /transl\_table=11  
 /product="protease"  
 /protein\_id="BAB55674.2"  
 /db\_xref="GI:205211155"  
 /translation="MRKKKVLFLSVLSAAAISTVALSNPSAGGARNFLDPKGIQTT  
 TDAGFSKQGTGAAPLVESENVLKPGLOKLETPANNKLHIQFNGPILLEETKQ  
 OLETKAKILDYIPDYAVIYVEGQDKSATSTIEHVESVEPLPIYRIDPOLPTKAS  
 ELVAVALDTQKKNKEVQLGIEQIAPALSNVDYLITAKPEYKWNVDVARGIVKADV  
 AQSSGLYGGQIVAVADTGLDTRNDSSNHEAFKGIITAIYALGTRNNANDTNGHT  
 HVAQSVLGNSTNKGMAPQANLVFQSDMSGGGLGSLNQLTFLFQVAGARIHTN  
 SWGAANGATYTSRNDYDVRKNDMTILFAAGNEGNGGTISAPGTAKNAITVGATE  
 NLRPSFGYADNINHVAQFSRSGPTDKGRIPKDVMAPGTFLSARSLAPDSSFVANH  
 DSKVYMGGTSMATPIVAGNVAOLRSHFVNKRGITTPKSLKLAALAGADIGLYPN  
 GNCQGVITLDSLVAVYVNESLSLTSOKATYSFTATAGKPLKISLVWSDAPASTTA  
 SVTLNDLDLVITAPNGCTQYVGNDFTSYNDNDGNNVENVFINAPQSGTITIEVQA  
 INVPVGFQTFSLAIVN"

ORIGIN  
 Alignment Scores:  
 Pred. No.: 1,31e-125 Length: 1923  
 Score: 2125.50 Matches: 406  
 Percent Similarity: 97.93% Conservative: 19  
 Best Local Similarity: 93.55% Mismatches: 8  
 Query Match: 94.36% Indels: 1

DB: 1 Gaps: 1  
US-09-985-689A-7 (1-433) x AB051423 (1-1923)

Qy 1 AsnAspValAlaArgGlyIleValValAlaAspValAlaGlnAsnAsnPhcGlyLeuTyr 20  
Db 619 AATGATGTTGGCGTGGAAATGTCACAGCGGATGTGGCTCAGAGCAGCTACGGGTTGTAT 678

Qy 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40  
Db 679 GGACAGGACAGATCGTAGCGTTGCCGATACAGGCTTGATACAGTGCATGACAGT 738

Qy 41 SerMetHisGluAlaPheArgGlyIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60  
Db 739 TCGATGATGAAGCTTCCGCGGAAATTAATCTCATTTATGATGATGGAGCGAGCAAT 798

Qy 61 AsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn--- 79  
Db 799 AATGCCAATGATACGAATGCTGATGCTAGCGATGTGGCTGCTCCGTATTAGGAACGGC 858

Qy 80 AlaThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 99  
Db 859 TCCACTAATAAGGAATGGCGCTCAGCGGAATCTAGTCTTCCCAATCTATCATGATGATG 918

Qy 100 GlyGlyGlyLeuGlyGlyLeuProAlaAsnLeuGlnThrLeuPheSerGlnAlaTyrSer 119  
Db 919 GGTGGGGACTTGGAGGACTACCTTCGAACTTGCACCTTATTCAGCCAGCATACAGT 978

Qy 120 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyrThr 139  
Db 979 GCTGTGTCAGAAATTCATAAACTCTCGGGAGCAGCAGTCAATGGCGCTTACACACA 1038

Qy 140 AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 159  
Db 1039 GATTCCAGAAATGTGGATGACTATGTGCGCAAAATGATATGACGATCTTTTCGTGCC 1098

Qy 160 GlyAsnGluGlyProGlySerGlyThrIleSerAlaProGlyThrAlaLysAlaIle 179  
Db 1099 GCGAATGAAGGACCGAAGCGGGAACCATCAGTCACCGAGCAGCATAAAAATGCAATA 1158

Qy 180 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 199  
Db 1159 ACAGTCGGAGCTACCGAAACCTCCGCCCAAGCTTTGGGTCTTATCGGCAATATCAAC 1218

Qy 200 HisValAlaGlnPheSerSerArgGlyProThrArgAspGlyArgIleLysProAspVal 219  
Db 1219 CATGTGGCACAGTCTCTTCACGTGGACCGCAAGGATGGAGCGATCAACCGGATGTC 1278

Qy 220 MetAlaProGlyThrThrIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 239  
Db 1279 ATGGCACCGGAACTTCATATCATCAGCAAGATCTCTTTCACCGGATTCCTCCTTC 1338

Qy 240 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProile 259  
Db 1339 TGGCGCAACATGACAGTAATATGATACATGCGGTGGACGTCTCATGGCTACACCGATC 1398

Qy 260 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyValThrPro 279  
Db 1399 GTTGTGGAACCGTGGCACAGCTTCGTGAGCATTTTGTGAAACACAGAGCATCACACA 1458

Qy 280 LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAspValGlyLeuGlyPhe 299  
Db 1459 AAGCCTTCCTCTATAAAGCGGACCTGATGTCGGTGCAGTGCATCGCCCTTGGCTAC 1518

Qy 300 ProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaPhe 319  
Db 1519 CCGAACCGTAACCAAGATGGGACGAGTGACATTCGATAAATCCCTGAACGTTGCCTAT 1578

Qy 320 ValAsnGluThrSerProLeuSerThrSerGlnLysAlaThrTyrSerPheThrAlaGln 339  
Db 1579 GTGAACGAGTCCAGTCTCTATCCACGCGCAAAAGCGGACGATCTCGTTTACTGCTACT 1638

Qy 340 AlaGlyLysProLeuLysIleSerLeuValTrpSerAspAlaProGlySerThrThrAla 359

Db 1639 GCCGCAAGCCTTTGAAAATCTCCCTGATGCTGTGATGCTCGGACACACTGCT 1698

Qy 360 SerLeuThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrLysTyr 379  
Db 1699 TCCGTAACGCTTGTCAATGATCTGACCTTGTCTATTACCGTCCAAATGGCACACAGTAT 1758

Qy 380 ValGlyAsnAspPheThrAlaProTyrAspAsnAsnTrpAspGlyArgAsnAsnValGlu 399  
Db 1759 GTAGAAATGACTTTTACTTCGCCATACATGATAACTGGGATGGCGCAATAACGTAGAA 1818

Qy 400 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrValGluValGlnAlaTyrAsn 419  
Db 1819 ATGTATTTATTTATATGACCAACAGCGGACGTATACATATGAGGTACAGGCTTATAAC 1878

Qy 420 ValProValSerProGlnThrPheSerLeuAlaIleValHis 433  
Db 1879 GTACCGGTTGGACACAGACCTTCTCGTTGGCAATTGGAAAT 1920

RESULT 6  
AR368117 1923 bp DNA linear PAT 12-SEP-2003  
LOCUS AR368117  
DEFINITION Sequence 5 from patent US 6376227.  
ACCESSION AR368117  
VERSION AR368117.1 GI:34601778  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 1923)  
AUTHORS Takaiwa,M., Okuda,M., Saeki,K., Kubota,H., Hitomi,J., Kageyama,Y.,  
Shikata,S. and Nomura,M.  
TITLE Alkaline protease  
JOURNAL Patent: US 6376227-A 5 23-APR-2002;  
FEATURES Location/Qualifiers  
source 1..1923  
/organism="unknown"  
/mol\_type="genomic DNA"

ORIGIN  
Alignment Scores: 1,31e-125 Length: 1923  
Pred. No.: 2125.50 Matches: 406  
Score: 2125.50  
Percent Similarity: 97.93% Conservative: 19  
Best Local Similarity: 93.55% Mismatches: 8  
Query Match: 94.38% Indels: 1  
DB: 6 Gaps: 1

US-09-985-689A-7 (1-433) x AR368117 (1-1923)

Qy 1 AsnAspValAlaArgGlyIleValValAlaAspValAlaGlnAsnAsnPhcGlyLeuTyr 20  
Db 619 AATGATGTTGGCGTGGAAATGTCACAGCGGATGTGGCTCAGAGCAGCTACGGGTTGTAT 678

Qy 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40  
Db 679 GGACAGGACAGATCGTAGCGTTGCCGATACAGGCTTGATACAGTGCATGACAGT 738

Qy 41 SerMetHisGluAlaPheArgGlyIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60  
Db 739 TCGATGATGAAGCTTCCGCGGAAATTAATCTCATTTATGATGATGGAGCGAGCAAT 798

Qy 61 AsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn--- 79  
Db 799 AATGCCAATGATACGAATGCTGATGCTAGCGATGTGGCTGCTCCGTATTAGGAACGGC 858

Qy 80 AlaThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 99  
Db 859 TCCACTAATAAGGAATGGCGCTCAGCGGAATCTAGTCTTCCCAATCTATCATGATGATG 918

Qy 100 GlyGlyGlyLeuGlyGlyLeuProAlaAsnLeuGlnThrLeuPheSerGlnAlaTyrSer 119  
Db 919 GGTGGGGACTTGGAGGACTACCTTCGAACTTGCACCTTATTCAGCCAGCATACAGT 978



Qy 120 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyThrThr 139  
 Db 979 GCTGTGTCAGAAATTCATCAAACTCCCTGGGAGCAGCAGTGAATGGGCTTACACACA 1038  
 Qy 140 AspSerArgAsnValAspAspTrpValArgLysAsnAspMetThrIleLeuPheAlaAla 159  
 Db 1039 GATTCAGAAATGTGGATGACTATGTGGCCAAAATGATATGACGATCTTTTCGCTGCC 1098  
 Qy 160 GlyAsnGlyProGlySerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 179  
 Db 1099 GGGAAATGAAGGACCGGACCGGAAACCATCAGTCAGCCAGGACAGCTAAAATGCAATA 1158  
 Qy 180 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTrpAlaAspAsnIleAsn 199  
 Db 1159 ACAGTCGGAGCTACGGAAAACCTCCGCCCAAGCTTTGGGTCTTATGGGACAATATCAAC 1218  
 Qy 200 HisValAlaGlnPheSerSerArgGlyProThrArgAspGlyArgIleLysProAspVal 219  
 Db 1219 CATGTGGCACAGTTCTCTTCACGTGGACCGCAAGAGGATGGACGGATCAAAACCGGATGC 1278  
 Qy 220 MetAlaProGlyThrTyrrIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 239  
 Db 1279 ATGGCACCAGGAACTCATATCATACAGCAAGATCTTCTTGTGACCGGATCTCTCTTC 1338  
 Qy 240 TrpAlaAsnHisAspSerLysTyrrAlaTyrrMetGlyGlyThrSerMetAlaThrProIle 259  
 Db 1339 TGGCGCAACCATGACAGTAAATATGCATACATGCGGTGGAACGTCCATGGCTACACCGATC 1398  
 Qy 260 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyValThrPro 279  
 Db 1399 GTTGTCTGGAACCGTGGCACAGCTTCGTGAGCATTTTGTGAAAACACAGAGGATCACACCA 1458  
 Qy 280 LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspValGlyLeuGlyPhe 299  
 Db 1459 AAGCCCTCTCTATTAAAGCGGCATGATGTCGCGGTGAGCTGACATCGGCTTGGCTAC 1518  
 Qy 300 ProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaPhe 319  
 Db 1519 CCGAACGGTAACCAAGGTTGGGAGCGAGTGACATTTGGATTAATCCTGTAAGCTTGCTAT 1578  
 Qy 320 ValAsnGluThrSerProLeuSerThrSerGlnLysAlaThrTyrrSerPheThrAlaGln 339  
 Db 1579 GTGAACGAGTCCAGTTCTCTATCCACCGCAGCATGATGCGGTGAGCTGACATCGGCTTACT 1638  
 Qy 340 AlaGlyLysProLeuLysIleSerLeuValTrpSerAspAlaProGlySerThrAlaAla 359  
 Db 1639 GCGGCAAGCCTTGAATCTCCCTGGTATGCTGATGCTGATGCTGCGGACGACAACTGCT 1698  
 Qy 360 SerLeuThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrLysTyrr 379  
 Db 1699 TCGTAACGCTTGTCAATGATCTGGACCTTGTCTATACCGCTTCCAAATGGCACACAGTAT 1758  
 Qy 380 ValGlyAsnAspPheThrAlaProTrpAspAsnAsnTrpAspGlyArgAsnValGlu 399  
 Db 1759 GTAGGAAATGACTTTACTTCGCCATACATGATACTGGATGGCGGCAATAAGCTAGAA 1816  
 Qy 400 AsnValPheIleAsnAlaProGlnSerGlyThrTyrrThrValGluValGlnAlaTyrrAsn 419  
 Db 1819 AATGTATTATTAAATGACCAAAAGCGGAGCGGTATACAAATGAGGTACAGGCTTATAAC 1878  
 Qy 420 ValProValSerProGlnThrPheSerLeuAlaIleValHis 433  
 Db 1879 GTACCGGTTGGACACAGACCTTCTCGTTGGCAATTTGGAT 1920

## RESULT 7

AR368118 LOCUS 1923 bp DNA linear PAT 12-SEP-2003  
 DEFINITION Sequence 7 from patent US 6376227.  
 AR368118 ACCESSION  
 AR368118.1 GI:34601779  
 KEYWORDS  
 Unknown.

## ORGANISM

Unknown.

Unclassified.

REFERENCE 1 (bases 1 to 1923)

AUTHORS Takaiwa,M., Okuda,M., Saeki,K., Kubota,H., Hitomi,J., Kageyama,Y., Shikata,S. and Nomura,M.

TITLE Alkaline protease

JOURNAL Patent: US 6376227-A 7 23-APR-2002;

FEATURES Location/Qualifiers

source

1..1923

/organism="unknown"

/mol\_type="genomic DNA"

## ORIGIN

## Alignment Scores:

Pred. No.: 2,71e-125 Length: 1923

Score: 2120.50 Matches: 405

Percent Similarity: 97.70% Conservative: 19

Best Local Similarity: 93.32% Mismatches: 9

Query Match: 94.16% Indels: 1

DB: 6 Gaps: 1

US-09-985-689A-7 (1-433) x AR368118 (1-1923)

Qy 1 AsnAspValAlaArgGlyIleValLysAlaAspValAlaGlnAsnAsnPheGlyLeuTyrr 20

Db 619 AATGATGTTGGCGTGGAAATGTCAACGGGATGTGGCTCAGACAGCTACGGGTTGTAT 678

Qy 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40

Db 679 GCACAAGACACATCGTAGCGTTGCCGATACAGGGCTTGATACAGGTGCGAATGACAGT 738

Qy 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrrAlaLeuGlyArgThrAsn 60

Db 739 TCGATGCGATGAAGCCCTCCGCGGAAATTTACTGCATTTATGATTTGGGACGACCAAT 798

Qy 61 AsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn--- 79

Db 799 AATGCCAATGATACGAATGGTCATGTCAGCATGTGGCTGCTCCGTATTAGGAAACGGC 858

Qy 80 AlaThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 99

Db 859 TCCACTAATAAAGGAATGGCGCTCAGCGAATCTAGTCTTCCANTCTATCATGGATAGC 918

Qy 100 GlyGlyGlyLeuGlyGlyLeuProAlaAsnLeuGlnThrLeuPheSerGlnAlaTyrrSer 119

Db 919 GGTGGGGAGCTTGGAGGACTACCTTCGAATCTGCAAACTTATTTCAGCAAGCATAAGT 978

Qy 120 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyrrThr 139

Db 979 GCTGTGGCAGAAATTCATACAACTCCCTGGGAGCAGCAGTGAATGGGCTTACACACA 1038

Qy 140 AspSerArgAsnValAspAspTrpValArgLysAsnAspMetThrIleLeuPheAlaAla 159

Db 1039 GATTCAGAAATGTGGATGACTATGTGGCGCAAAATGATATGACGATCTTTTCGCTGCC 1098

Qy 160 GlyAsnGlyProGlySerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 179

Db 1099 GGGAAATGAAGGACCGAAGCGCGGACCATCATGTCAGCCAGCAGCAGCTTAAATGCAATA 1158

Qy 180 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTrpAlaAspAsnIleAsn 199

Db 1159 ACAGTCGGAGCTACGGAAAACCTCCGCCCAAGCTTTGGGTCTTATGGGACAATATCAAC 1218

Qy 200 HisValAlaGlnPheSerSerArgGlyProThrArgAspGlyArgIleLysProAspVal 219

Db 1219 CATGTGGCACAGTTCTCTTCACGTGGACCGCAAGAGGATGGACGGATCAAAACCGGATGC 1278

Qy 220 MetAlaProGlyThrTyrrIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 239

Db 1279 ATGGCACCAGGAAACGTTCATATCATACAGCAAGATCTTCTTGTGACCGGATCTCTCTTC 1338

Qy 240 TrpAlaAsnHisAspSerLysTyrrAlaTyrrMetGlyGlyThrSerMetAlaThrProIle 259

```
Db 1339 TGGCGGACCATGACAGTAATATGATCATCATCGGTGGGAACGTCCTCATGCTACACCGATC 1398
Qy 260 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValIysAsnArgGlyValThrPro 279
Db 1399 GTTGCTGGAACCTGGGCACAGCTTCGTGAGCATTTTGTGAAAAACAGAGGCATCACACCA 1458
Qy 280 LysProSerLeuLeuIysAlaAlaLeuIleAlaGlyValAlaAlaAspValGlyLeuGlyPhe 299
Db 1459 AAGCTTTCTCTATTAAAGCGGCACTGATTGCGGTTGCAGTCACTGACCTGGCTTGGCTAC 1518
Qy 300 ProAsnGlyAsnGlnGlyTTPGlyArgValThrLeuAspIysSerLeuAsnValAlaPhe 319
Db 1519 CCGAACGCTTAACCAAGGATGGGACGAGTGACATTTGCAATATCCCTGGAACGTTGCTAT 1578
Qy 320 ValAsnGluThrSerProLeuSerThrSerGlnIysAlaThrIysSerPheThrAlaGln 339
Db 1579 GTGAACGAGTCCAGTCTCTATCCACGAGCAAAAGCGAGTACTGTTTACGTACT 1638
Qy 340 AlaGlyLysProLeuIysIleSerLeuValIrsSerAspAlaProGlySerThrAla 359
Db 1639 GCGGCAAGCCTTTGAAATCTCCCTGGTATGCTGATGCCCTGCGAGCACAACTGCT 1698
Qy 360 SerLeuThrLeuValAsnAspLeuAlaValIleThrAlaProAsnGlyThrIysTyr 379
Db 1699 TCCGTAAAGCTTGCAATGATCTGGACCTGTCTATTACCGCTCCAAATGGCACACAGTAT 1758
Qy 380 ValGlyAsnAspPheThrAlaProTyrAspAsnAsnTrpAspGlyArgAsnAsnValGlu 399
Db 1759 GTAGGAATGACTTTACTTCCCATACATGATGAATGGGATGGCCGCAATACGTAGAA 1818
Qy 400 AsnValPheIleAsnAlaProGlnSerGlyThrIysThrValGluValGlnAlaTyrAsn 419
Db 1819 AATGTATTATTATATGACCAAGCGGACGATATCAATGAATGAATGACAGGCTTATAAC 1878
Qy 420 ValProValSerProGlnThrPheSerLeuAlaIleValHis 433
Db 1879 GTACCGTGGACCAACAACTTCTCGTTGGCAATTGTGAAT 1920

RESULT 8
AB046406 1302 bp DNA linear BCT 23-JAN-2001
LOCUS Bacillus sp. NV1 PROBE gene for protease, partial cds.
DEFINITION Bacillus sp. NV1 PROBE gene for protease, partial cds.
ACCESSION AB046406
VERSION AB046406.1 GI:12381944
KEYWORDS
SOURCE Bacillus sp. NV1
ORGANISM Bacillus sp. NV1
REFERENCE 1 (sites)
AUTHORS Saeki,K., Okuda,M., Hatada,Y., Kobayashi,T., Ito,S., Takami,H. and
Horikoshi,K.
TITLE Novel oxidatively stable subtilisin-like serine proteases from
alkaliphilic Bacillus spp.: enzymatic properties, sequences, and
evolutionary relationships
JOURNAL Biochem. Biophys. Res. Commun. 279 (2), 313-319 (2000)
MEDLINE 20568675
PUBMED 11118284
REFERENCE 2 (bases 1 to 1302)
AUTHORS Saeki,K.
TITLE Direct Submission
JOURNAL Submitted (20-JUL-2000) Katsuhisa Saeki, Kao corporation,
Biological Science Laboratory; Ichikaimachi Akabane 2606, Haga,
Tochigi 321-3497, Japan (E-mail:387185@kasanet.kao.co.jp,
Tel:81-285-68-7400, Fax:81-285-68-7403)
FEATURES
source
1..1302
/organism="Bacillus sp. NV1"
/mol_type="genomic DNA"
/strain="NV1"
/db_xref="taxon:133781"
1..1302
/gene="PROE"
<1..1302
gene
1..1302
CDS
1..1302
```

```

/gene="PROE"
/codon_start=1
/transl_table=11
/proteome="protease"
/protein_id="BAB21269.1"
/db_xref="GI:12381945"
/translation="NDVARGIVKADVAQSSYGLYGGQVVAVADTGLDGRNDSMEH
AFPRKITAIVAGLRTNNDPNHGHTVAGSVLNGTSGNKGMAPQANLVFQSVNDSNG
AGLGLPNVSTLFSQAYSAQARIHNSWGAIPVNGAYTTDSRNVDVYRKNDMAVLFPA
GNEPNGGTTISAPCTAKNAITVGATNLRPSFGSYADINIHVAQFSFSGRPTKGRIFKP
DMPAPGPTIISARSLAPDPSFWANHDSKYAVMGTSMTAPIVAGNVQALREHFIKNR
GYTPKPSLLKALAGATIDGLGYPSNGQNGRVTLDKSLNVAFNVTSSISTNOKAT
YSFTAQSGKPKLISLVWSDPASTSASVTLLVDLIDLVTITAPNGTKYVGNDEFTAPYDNN
WDGRNNVENFENAPQSGTYTVEQAYNVPGQPAFSLAIVN"
ORIGIN
Alignment Scores:
Pred. No.: 3e-125 Length: 1302
Score: 2116.50 Matches: 400
Percent Similarity: 97.00% Conservatives: 21
Best Local Similarity: 92.17% Mismatches: 12
Query Match: 93.98% Indels: 1
DB: 1 Gaps: 1
US-09-985-689A-7 (1-433) x AB046406 (1-1302)
Qy 1 AsnAspValAlaArgGlyIleValIysAlaAspValAlaGlnAsnAsnPheGlyLeuTyr 20
Db 1 AACGATGTTGCCAGAGGCAATTGTAAGCCGATGTTGCCAGAGCATGTATGTTTATAT 60
Qy 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 61 GGGCAAGGCGCAAGTGGTTCAGTACGGATACCGACTGGATACAGGCGGTAATCACAGC 120
Qy 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
Db 121 TCGATGCATGAAGCGTTCGAGGAAAGATTACCGCGATATATGCCCTTGGAGAACAAAC 180
Qy 61 AsnAlaAspAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnAla 80
Db 181 AACGCCAATGATCCAAACGACACGGGACGCAATGTTGCCGATCTGTTTAGGAAACGGT 240
Qy 81 Thr---AsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 99
Db 241 ACTTCGAATAAAGGGATGCTCCACAAGCTAACTTAGTTTCCAACTCTGTTATGACACG 300
Qy 100 GlyGlyGlyLeuGlyGlyLeuProAlaAsnLeuGlnThrLeuPheSerGlnAlaTyrSer 119
Db 301 AATGGCGGACTTGGCGGACTGCTCCAAATGAATGATCATTTATTCAGCCAGCATATAGT 360
Qy 120 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyrThrThr 139
Db 361 GCCGGTGCCAGAAATCCATACGAACTCATGGGAGCGCTGTAAACGAGCGTACACTACT 420
Qy 140 AspSerArgAsnValAspAspTyrValAlaGlyAsnAspMetThrIleLeuPheAlaAla 159
Db 421 GATTCCAGAAACGTCAGCACTATGTTTCGTAAATAATGATAGCGGTTCTTTTTCAGCG 480
Qy 160 GlyAsnGluGlyProGlySerGlyThrIleSerAlaProGlyThrAlaIysAsnAlaIle 179
Db 481 GGTAAACGAGGCGGATGCGGCGCACATCAGTGTCTCTGTTACCGCAAGAAATGCTATC 540
Qy 180 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspIleAsn 199
Db 541 ACAGTAGGGGCAACAGAAAACCTGCGCCCAAGCTTTGGATCTTTATGCTGACAACTCAAT 600
Qy 200 HisValAlaGlnPheSerSerArgGlyProThrArgAspGlyArgIleLysProAspVal 219
Db 601 CATGTACACAGTTTTCCTCCCGGACCTTACAAAGATGGAGCTATCAAAACGGACGTA 660
Qy 220 MetAlaProGlyThrTyrIleLeuSerAlaArgSerSerIleuAlaProAspSerPhe 239
Db 661 ATGGCACCAAGAACATTTATTTTATCGGCAAGATCTCTTTTGGCTCCGGACTCCTCATTC 720
```

QY 240 TtpAlaAsnHisAspSerIysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProile 259  
 DB 721 TGGGCAAAACCATGACAGCAATATGCTTATATGGGTGGAACATCATCGGACACCGGAT 780  
 QY 260 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnAspGlyValThrPro 279  
 DB 781 GTAGCTGGTAAAGTTCACAGTACGTAAACATTCATCAAAACAGAGMAATCACCTCT 840  
 QY 280 LysProSerLeuLeuLysAlaAlaLeuLeuAlaGlyAlaAlaAspValGlyLeuGlyPhe 299  
 DB 841 AAACCATCTCTGCTGAAAGCAGCTCTTATTCGCGAGCAACTGATGATCGGTCTTGGCTAT 900  
 QY 300 ProAsnGlyAsnGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsnValAlaPhe 319  
 DB 901 CCGAGTGGAAACCAAGGATGGGAAGAGTAACATTTGGACAGTCTTATGATGCTTTC 960  
 QY 320 ValAsnGluThrSerProLeuSerThrSerGlnLysAlaThrTyrSerPheThrAlaGln 339  
 DB 961 GTAATGAAACCAAGCTCTTATCTACTAACCACCAAGGCTAGCTATTCATTACTGCAAA 1020  
 QY 340 AlaGlyLysProLeuLysIleSerLeuValTyrSerAspAlaProGlySerThrAla 359  
 DB 1021 TCAGGCARACCTTTGAGATTCATTTGTTGGTCTGATGCCCGCAAGTACTTCGCCA 1080  
 QY 360 SerLeuThrLeuValAsnAspLeuAspLeuValThrAlaProAsnGlyThrLysTyr 379  
 DB 1081 TCGGTATCATTTGGTGAATGATCTGGATCTGGTGTATACAGCTCCAAATGGAACAAAGTAT 1140  
 QY 380 ValGlyAsnAspPheThrAlaProTyrAspAsnAsnTyrAspGlyArgAsnValGlu 399  
 DB 1141 GTTGGAAACGACTTTACTGCTCCCTATGATATTAATTAACCTGGATGACCTACATGTAGAG 1200  
 QY 400 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrValGluValGlnAlaTyrAsn 419  
 DB 1201 AACGTGTTTATCAATGCTCCGCAAGCGGAACGTATACAGTTGAGGTACAGGCTTACAAT 1260  
 QY 420 ValProValSerProGlnThrPheSerLeuAlaIleValHis 433  
 DB 1261 GTTCCACAGGCGCCGAGCGCTTTCTTTGGCTATTGTGAC 1302

RESULT 9  
 AB046403  
 LOCUS  
 DEFINITION  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 REFERENCE  
 AUTHORS

AB046403 1920 bp DNA linear BCT 10-MAY-2002  
 Bacillus sp. 9860 PROA gene for protease, complete cds.  
 AB046403  
 AB046403.2 GI:20521152  
 Bacillus sp. 9860  
 Bacillus sp. 9860  
 Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
 1 Saeki, K., Okuda, M., Hatada, Y., Kobayashi, T., Ito, S., Takami, H. and  
 Horikoshi, K.  
 Novel oxidatively stable subtilisin-like serine proteases from  
 alkaliphilic Bacillus spp.: enzymatic properties, sequences, and  
 evolutionary relationships  
 Biochem. Biophys. Res. Commun. 279 (2), 313-319 (2000)  
 20568675  
 11118284  
 2 (bases 1 to 1920)  
 Saeki, K.  
 Direct Submission  
 Submitted (20-JUL-2000) Katsuhisa Saeki, Kao corporation,  
 Biological Science Laboratory, Ichikaimachi Akabane 2606, Haga,  
 Tochigi 321-3497, Japan (E-mail:387.185@kamatanet.kao.co.jp,  
 Tel:81-285-68-7400, Fax:81-285-68-7403)  
 On May 9, 2002 this sequence version replaced gi:12381938.  
 Location/Qualifiers  
 1..1920  
 /organism="Bacillus sp. 9860"  
 /mol\_type="genomic DNA"  
 /strain="9860"

JOURNAL  
 MEDLINE  
 PUBMED  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL

COMMENT  
 FEATURES  
 source

/db\_xref="taxon:133778"  
 1..1920  
 /gene="PROA"  
 1..1920  
 /gene="PROA"  
 /codon\_start=1  
 /transl\_table=11  
 /product="protease"  
 /protein\_id="BAB21266.2"  
 /db\_xref="GI:20521153"  
 /translation="MRKKVFLSVLSAAAILSTVALNPPSAGDARTFDLDFKGIQTWT  
 DVSGFSKQRTGAAAFVSESNVGLKLLKLETVANNKHLVQFNGPILBETKQK  
 LETTGAKLDVDPYAVIYEGDGVSKVSEIEHVSVEPYLPKYKDPLFKGAST  
 LVKALADTKNNKEVOLGIEIAOVASNDVHYLTAKPEYKVMNDVARGIVKAQVA  
 OSSGLVGGQGLVAVADTGLDGTGNSMEEAFRGKITALYALGRNNDNTNGHGH  
 VAGSVNGATYKDSNRVDYVYKNDMTILFAAGNERPNGGTISAPGAKNAITVGATNS  
 WGSVNGAYTTSRVDYVYKNDMTILFAAGNERPNGGTISAPGAKNAITVGATNS  
 LRPFSGYANINHVAFSGRPTKGRIPKDYNAFPTVLSARSLADSPFMANHD  
 SKYAMGTSMATIVAGNVQALREHPKVRGITPKPSLLKALIAAGADVGLGYGNG  
 NQGMGRVTLQKSLNVAVYNSSSALSTQKATYTTATAGKPLKISLWSDAPASTAS  
 VTLVNDLDELITAPNGTRYVNDPFSAPFDNNWDGRNVENVFINSFQSGTYTTEVQAY  
 NVPVGPQNFSIAIVN"

ORIGIN  
 Alignment Scores:  
 Pred. No.: 1.16e-124 Length: 1920  
 Score: 2110.50 Matches: 404  
 Percent Similarity: 97.24% Conservative: 18  
 Best Local Similarity: 93.09% Mismatches: 11  
 Query Match: 93.72% Indels: 1  
 DB: 1 Gaps: 1

US-09-985-689A-7 (1-433) x AB046403 (1-1920)

QY 1 AsnAspValAlaArgGlyIleValLysAlaAspValAlaGlnAsnAsnAspPheGlyLeuTyr 20  
 DB 616 AATGATGCGGACAGAGTATGTCAAAGCGGATGTGGCAGAGGAGTACGCTTTGTAT 675  
 QY 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40  
 DB 676 GGACAGGCCAGATTGCGGAGTTGCGGATGCTGGATTTGATACAGGAAGAACGACAGT 735  
 QY 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60  
 DB 736 TCGATGATGAAGCTTCCGCGGTAAATAACAGCACTATATGCACTGGGTGCGACGAT 795  
 QY 61 AsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn--- 79  
 DB 796 AATGCGAATGATACGAACGGTCAATGGTACCCTATGCGAGGTCGGTATTAGGAATGGC 855  
 QY 80 AlaThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 99  
 DB 856 GCAACGAATAAAGCAATGGCACCTCAAGCGAATCTGGTTTTCATCCATCATGATAGC 915  
 QY 100 GlyGlyGlyLeuGlyGlyLeuProAlaAsnLeuGlnThrLeuPheSerGlnAlaTyrSer 119  
 DB 916 AGTGTGGGCTTGAGGCGCTTCCCTTCCAATCTGCAACCTTATTCAGCCCAAGCATTCAGT 975  
 QY 120 AlaGlyAlaArgIleHisThrAsnSerTyrGlyValProValAsnGlyAlaTyrThr 139  
 DB 976 GCAGGTGCCAGATTATCAAACTCTCTGGGGGCGAGCGGTGAATGGGGCTACACGACA 1035  
 QY 140 AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 159  
 DB 1036 CATTCAGAAATATGGATGACTATGTAAAGGAAAATATATGACGATCTTTTCGGCGCT 1095  
 QY 160 GlyAsnGluGlyProGlySerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 179  
 DB 1096 GGGAAATGAAGGCCCAACGGCGGTACCATCAGTGCACCTGGTACGGCTAAAACGCCATA 1155  
 QY 180 ThrValGlyAlaThrGluAsnLeuAspGlyProSerPheGlySerTyrAlaAspAsnIleAsn 199  
 DB 1156 ACAGTGGCGGCACACCGAAACCTGCGGTCCAAGCTTCGGTTCCTATGACAGATAATATTAC 1215



```
Db 1576 GTGAACGATCCAGTGCCTTATCAACTAGCCAAAAGGACATATACCTTTACTGCAACG 1635
Qy 340 AlGlylySPProLeuylsIleSerLeuValTrpSerAspAlaProGlySerThrThrAla 359
Db 1636 GCGGGCAAGCATGAAATCTCCCTGGTATGTCGATGCGCCCTCAAGACATCTGCT 1695
Qy 360 SerLeuThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrIleTyr 379
Db 1696 TCTGTAACTTGGTCAATGATTTGGATTTGGTCAATACAGCACCAAAACGAAACAGATAT 1755
Qy 380 ValGlyAsnAspPheThrAlaProTyrAspAsnAsnTrpAspGlyArgAsnValGlu 399
Db 1756 GTGCGGAATGACTTCTCAGACACCATTTGCACTAATCTGGATGGGCGCAATAACGTAGAA 1815
Qy 400 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrValGluValGlnAlaTyrAsn 419
Db 1816 AATGTATTTAATATTCGCCCAAGTGGCAATATACCATTTGAGTGGCAAGCATATAAT 1875
Qy 420 ValProValSerProGlnThrPheSerLeuAlaIleValHis 433
Db 1876 GTGCGGGTGGACCAAACTTCTCGTTGGCAATTTGTGAAC 1917

RESULT 11
AB046405 1299 bp DNA linear BCT 23-JAN-2001
LOCUS Bacillus sp. SD521 PROD gene for protease, partial cds.
DEFINITION Bacillus sp. SD521
ACCESSION AB046405
VERSION AB046405.1 GI:12381942
KEYWORDS
SOURCE Bacillus sp. SD521
ORGANISM Bacillus sp. SD521
Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
REFERENCE 1 (sites)
AUTHORS Saeki,K., Okuda,M., Hatada,Y., Kobayashi,T., Ito,S., Takami,H. and
Horiuchi,K.
TITLE Novel oxidatively stable subtilisin-like serine proteases from
alkaliphilic Bacillus spp.: enzymatic properties, sequences, and
evolutionary relationships
JOURNAL Biochem. Biophys. Res. Commun. 279 (2), 313-319 (2000)
MEDLINE 20568675
PUBMED 11118284
REFERENCE 2 (bases 1 to 1299)
AUTHORS Saeki,K.
DIRECT SUBMISSION
TITLE Submitted (20-JUL-2000) Katsuhisa Saeki, Kao corporation,
JOURNAL Biological Science Laboratory, Ichikaimachi Akabane 2606, Haga,
Tochigi 321-3497, Japan (E-mail:387185@kasanet.kao.co.jp,
Tel:81-285-68-7400, Fax:81-285-68-7403)
FEATURES
Location/Qualifiers
1..1299
/organism="Bacillus sp. SD521"
/mol_type="genomic DNA"
/strain="SD521"
/db_xref="taxon:133780"
1..1299
/gene="PROD"
<1..>1299
/gene="PROD"
/codon_start=1
/transl_table=11
/product="protease"
/protein_id="BAB21268.1"
/db_xref="GI:12381943"
/translation="NDVARGIVKADYVAQNNGYLYGGQVAVADVADTGLDTRNDSSMHE
AFRGKITALYALGRNTNANDPNGHTHVAGSVLGNALNKGMAPQANLVFQSIIMPSSGG
LGLPSNLNTLSQAKNAGARLHTNSWGA PVNGAYTANSROVDEYVRNNDMTLVFPAAG
NKGPSGTISAGTKAATVATENATNSRPSFGSLADNPNHIAQSSRCATRDGRFKPD
VTAPGFFILSAKSLAPDSSFWANTNSKTYAMGTSMTATPIVAGNVAQIRHEFIKRG
ITPKSLIKALALAGATLDFGPGDQGWRTLDKSLNVAYNVEALATGQATY
SFOAQAGKPLSLVWTDAPGSTTASLYLNDLDLVITAPNGQKTVGNDPSYPYDNNV
DGRNNVENVFINAPQSGTVTIEVQAVNNVPSQPRFSLAIH"
```

```
Alignment Scores:
Pred. No.: 1,16e-120 Length: 1299
Score: 2044.00 Matches: 386
Percent Similarity: 94.69% Conservative: 24
Best Local Similarity: 89.15% Mismatches: 23
Query Match: 90.76% Indels: 0
DB: 1 Gaps: 0

US-09-985-689A-7 (1-433) x AB046405 (1-1299)

Qy 1 AsnAspValAlaAtrGlyIleValIysAlaAspValAlaGlnAsnAsnPheGlyLeuTyr 20
Db 1 AATGTATGTAGCAAGAGGATAGTAAAGCAGACGTTGTCACAAAACAATATCGGACTATAT 60
Qy 21 GlyGlnGlyGlnIleValAlaAlaAspThrGlyLeuAspThrGlyIleValIysAsnSer 40
Db 61 GCGAAGGTCAAGTATGTCAGTAGCGGATAGCGGTTTAGATACAGGTCGTAAACATAGT 120
Qy 41 SerMetHisGluAlaPheArgGlyIleThrAlaLeuTyrAlaLeuGlyIleValIysAsn 60
Db 121 TCTATGTCATGAACATTCGTGGGAAATATACAGCTCTTTACGCGTTAGGAAGAACTAAC 180
Qy 61 AsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValIleGlyAsnAla 80
Db 181 AATGCAATATGATCCGAATGGGATGTCATGTCATGTCATGTCATGTCATGTCATGTCAT 240
Qy 81 ThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSerGly 100
Db 241 TTAATAAAGGAATGCTCCGACAGCTAATCTAGCTCTTCCATCTATCTATGATAGCAGC 300
Qy 101 GlyGlyLeuGlyGlyLeuProAlaAsnLeuGlnThrLeuPheSerGlnAlaTyrSerAla 120
Db 301 GCGAGGATTAGTGGATTACCATCGAATTTGAATAGCTTATTTAGTCAAGCTTGGAAATGT 360
Qy 121 GlyAlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyValAlaTyrThrAsp 140
Db 361 GCGGCTAGAAATTCATACTAATCTTGGGTGCTCCAGTAAATGGAGCGTACACTGCTAAC 420
Qy 141 SerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaGly 160
Db 421 TCGAGACAAGTGGATGAGTATGTTGAAATAATATGATATGACGCTACTTTTTCAGCAGGT 480
Qy 161 AsnGluGlyProGlySerGlyThrIleSerAlaProGlyThrAlaIleValAlaIleThr 180
Db 481 AATGAAGAGTCTTAATTCAGGAACAATTAGTGTCTCCAGGCACAGCGGAAAAATGCCATTAG 540
Qy 181 ValGlyValThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsnHis 200
Db 541 GTCGGCGCAACGGAACAACTATCGCCCGAGCTTCGTTTCATTAGCAGATAACCCCAATCAT 600
Qy 201 ValAlaGlnPheSerSerArgGlyProThrArgAspGlyArgIleLysProAspValMet 220
Db 601 ATCGCACATTTTTCATCAGAGAGAGCTACGAGAGATGGAAGATTAACCGACGTAACA 660
Qy 221 AlaProGlyThrTyrIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrp 240
Db 661 GCTCTCTGAAACATTTATTTTATGACACGTTCTTCTTCTAGCCCGACACTCTTCGTTTGG 720
Qy 241 AlaAsnHisAspSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProIleVal 260
Db 721 GCGAATTTAACAAGTAAAGTATGCGTACATCGGCGGTACTCTTATGCGGACACCTATAGTT 780
Qy 261 AlaGlyAsnValAlaGlnLeuArgGluHisPheValIleAsnArgGlyValThrProLys 280
Db 781 GCGGGGAATGTCGCGCAATTTACGCGAGCATTTTATATAAATAATAGAGAAATACACCTAAA 840
Qy 281 ProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspValGlyLeuGlyPhePro 300
Db 841 CCTTCTCTAATAAAGCTGCACCTATCGCTGGGCTACTGATGTTGGTTGGTATAGGATATCCA 900
Qy 301 AsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaPheVal 320
Db 901 AGTGGTGACCAAGGCTCGGGGCGGTGTACTCTAGATAAATCGTTAAATATGCGTATGTC 960
```

Qy 321 AsnGluThrSerProLeuSerThrSerGlnlyAlaThrTyrSerPheThrAlaGlnAla 340  
Db 961 AATGAACCAATCGCATTTAGCAGCAACAAAAGCAACGTTATTCGTTCACAGCACAAGCG 1020  
Qy 341 GlyIysProLeuLysIleSerLeuValTyrSerAspAlaProGlySerThrAlaSer 360  
Db 1021 GGTAAACCTTTAAANAATCTGTTAGTATGACAGATGACCTGGAAGTACAACCTGCATCT 1080  
Qy 361 LeuThrLeuValAsnAspLeuAlaProGlnSerGlyThrValGluValGlnAlaTyrAsnVal 380  
Db 1081 TATACACTAGTTATGATTTAGATCTAGTTATTTACTCTCCGATCGCAAAAATATGTA 1140  
Qy 381 GlyAsnAspPheThrAlaProTyrAspAsnAsnTyrAspGlyArgAsnValGluAsn 400  
Db 1141 GGAATAGATTTTATGTTATCTTATGATTAATACCTGGATCGTCCCAACAATGTTGAGAC 1200  
Qy 401 ValPheLeuAsnAlaProGlnSerGlyThrTyrThrValGluValGlnAlaTyrAsnVal 420  
Db 1201 GTATTTATAACCGTCCGCAATCTGGAACGTATACAAATTCAGGTTCAAGCGTATAATGTA 1260  
Qy 421 ProValSerProGlnThrPheSerLeuAlaIleValHis 433  
Db 1261 CCACTGTGCCACACGCTTCTACTAGTATCTGATCAT 1299

RESULT 12  
LOCUS AB046402 1299 bp DNA linear BCT 16-JUN-2001  
DEFINITION Bacillus sp. D6 PROA gene for protease, partial cds.  
ACCESSION AB046402.1 GI:12381936  
VERSION AB046402.1  
KEYWORDS  
SOURCE  
ORGANISM Bacillus sp. D6  
Bacillus sp. D6  
Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
1 (sites)  
Saeki, K.; Okuda, M.; Harada, Y.; Kobayashi, T.; Ito, S.; Takami, H. and  
Horikoshi, K.  
TITLE Novel oxidatively stable subtilisin-like serine proteases from  
alkaliphilic Bacillus spp.: enzymatic properties, sequences, and  
evolutionary relationships  
Biochem. Biophys. Res. Commun. 279 (2), 313-319 (2000)  
20568675  
11118284  
2 (bases 1 to 1299)  
Saeki, K.  
Direct Submission  
Submitted (20-JUL-2000) Katsuhisa Saeki, Kao corporation,  
Biological Science Laboratory; Ichikaimachi Akabane 2606, Haga,  
Tochigi 321-3497, Japan (E-mail:367185@kasanet.kao.co.jp,  
Tel:81-285-68-7400, Fax:81-285-68-7403)  
LOCATION/Qualifiers  
1. 1299  
/organism="Bacillus sp. D6"  
/mol\_type="genomic DNA"  
/strain="D6"  
/db\_xref="taxon:127889"  
1. 1299  
/gene="PROA"  
<1..>1299  
/codon\_start=1  
/gene="PROA"  
/product="protease"  
/transl\_table=11  
/protein\_id="BAB21265.1"  
/db\_xref="GI:12381937"  
/translation="NDVARGIVKADVAQNNYGLYKQGVAVADTGLTGRDSSMHE  
AFRGKITVALYALGRTNANDPNHGHTVAGSVLGNALNKGMAPQANLVFQSIMSSGG  
LGLPSENLTFLSQWAGARLHTNSWGAPVNGVAYTANRSQVDEYVRNNDMTLFAAG  
NKGPNSTISAPGTAKNAITVGATENYRPFSGSIADNPNIHQFSSRGATRDGRKPD  
VTAPGPFILSSAPLAPDSSFWANYSKYVMGTSMTAPIVAGNVAOLREHFIKNRG  
ITPKSLIKALTAGTDVLGVPSGDQGWGRVTLDKSLNVAYNEATLTGOKATY  
SFQTQAGKPEKISLWTDAPGTSATSYTLVNDLVLITAPNGQKIVGNDPFPYDNNW

FEATURES  
source

gene

CDS

## ORIGIN

Alignment Scores:  
Pred. No.: 2,08e-120 Length: 1299  
Score: 2040.00 Matches: 385  
Percent Similarity: 94.46% Conservatives: 24  
Best Local Similarity: 88.91% Mismatches: 24  
Query Match: 90.59% Indels: 0  
DB: 1 Gaps: 0  
US-09-985-689A-7 (1-433) x AB046402 (1-1299)  
Qy 1 AsnAspValAlaAaArgGlyIleValIysAlaAspValAlaGlnAsnAsnPheGlyLeuTyr 20  
Db 1 AATGATGTAGCAAGAGGAATAGTAAAGCAGACGCTTGCAAAAACAAATATCGGACTATAT 60  
Qy 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40  
Db 61 GGCAAGGTCTAGTGTAGTTCAGATGCGGATAGCGGTTTAGATACAGGTCTGACGATAGT 120  
Qy 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60  
Db 121 TCTATGCGATGAGCATTCGGTGGGAAATTTACAGCTCTTTACGGCTTAGGAAGAACTAAC 180  
Qy 61 AsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnAla 80  
Db 181 AATGCAAAATGATCCCAATGGGATGCTAGCTAGTGTCTCTGCTCTGCTTGGTATGCT 240  
Qy 81 ThrAsnLysGlyWetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSerGly 100  
Db 241 TTAATAAAGGAATGGCTCCGCAAGCTAACTTAGTCTTCCAATCTATTATGATAGCAGC 300  
Qy 101 GlyGlyLeuGlyGlyLeuProAlaAsnLeuGlnThrLeuPheSerGlnAlaTyrSerAla 120  
Db 301 GGAGGATTAGGAGGATTACCATCGAATTTAAATACGTTATTATTAGTCAAGCTTGGATGCT 360  
Qy 121 GlyAlaAlaIleHisThrAsnSerTyrGlyAlaProValAsnGlyValaTyrThrAsp 140  
Db 361 GGCGCTAGAAATTCATACTAACCTCTGGGGGGCCCCAGTAATATGGAGCGTACACTGCTAAC 420  
Qy 141 SerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaGly 160  
Db 421 TCGAGCAAGTGGATGAGTATGTTTCGAAACAATGATATGACGGTACTTTTTCAGCTGGA 480  
Qy 161 AsnGlyGlyProGlySerGlyThrIleSerAlaProGlyThrAlaIysAsnAlaIleThr 180  
Db 481 AATGAAGGACCTTAACCTCTGGAACAATTAGCGCTCCAGGACAGCGAATAATGCCATTACG 540  
Qy 181 ValGlyValaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsnHis 200  
Db 541 GTCGGCGCAACGGAAACCTACCGCAAGTTTGTTCATTTGTCAGATTAACCTTAATCAT 600  
Qy 201 ValAlaGlnPheSerSerArgGlyProThrArgAspGlyValArgIleLysProAspValMet 220  
Db 601 ATCGCACAATTTTTCATCGAGAGAGGTACGAGAGATGGAGCAATTAACACGACGCTAAC 660  
Qy 221 AlaProGlyThrTyrIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTyr 240  
Db 661 GCTCTCGACATTTATATATTATCATGACGCGCTCTCTTTAGCACGACCTCTCGTTTGG 720  
Qy 241 AlaAsnHisAspSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProIleVal 260  
Db 721 GCGAATTTATACAGTAAATGCTATGCTATATGGGCGGTACCTCTATGCGGACACCTATGTT 780  
Qy 261 AlaGlyAsnValAlaGlnLeuArgGluHisPheValIysAsnArgGlyValThrProLys 280  
Db 781 GCGGGGGAATGTCGGCAATTTACCGGAGCAATTTTATAAAAAATAGAGAAATTACACCTAAA 840  
Qy 281 ProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspValGlyLeuGlyPhePro 300  
Db 841 CCTTCTCTTAATAAAGTGCACCTATTCGCTGGGGGCTACTGATGTTGCTTAGGATATCCA 900



QY 301 AsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaPheVal 320  
 Db 901 AGTGGTACCAAGGCTGGGGCGTGTACTTTAGATAAATCTTAAATGTAGCGTATGTC 960  
 QY 321 AsnGlyThrSerProLeuSerThrSerGlnLysAlaThrTySerPheThrAlaGlnAla 340  
 Db 961 AATGAAGCAACTGCATTAACAACAGACAAAAGCAACGTATTCGTTCCAAACGCAAGCG 1020  
 QY 341 GlyLysProLeuLysIleSerLeuValTrpSerAspAlaProGlySerThrThrAlaSer 360  
 Db 1021 GGTAAACCACTTAAATCTCGTTAGTATGGACAGATGCACCTGGAGTACACAGCATCT 1080  
 QY 361 LeuThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrLysTyVal 380  
 Db 1081 TATACACTAGTTAATCAATTAGTCTAGTATTAATCTGCTCCGAATGGACAAAATATGTA 1140  
 QY 381 GlyAsnAspPheThrAlaProTyAspAsnAsnTrpAspGlyArgAsnValGluAsn 400  
 Db 1141 GGTAAATGTTTGTATCTTATGATATAATTTGGATGGTTCGCAACAATGTGAGAAC 1200  
 QY 401 ValPheIleAsnAlaProGlnSerGlyThrTyThrValGluValGlnAlaTyAsnVal 420  
 Db 1201 GTATTATTAACCTCCGAATCTGGACAGTATACAATTTGAGGTTCAAGCGTATAACGTT 1260  
 QY 421 ProValSerProGlnThrPheSerLeuAlaIleValHis 433  
 Db 1261 CCATCTGGACCACAGCGTTTTCACCTAGCTAGCTATCGTACAT 1299

## RESULT 13

AB046404  
 LOCUS  
 DEFINITION  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
 1 (sites)  
 Saeki, K., Okuda, M., Hatada, Y., Kobayashi, T., Ito, S., Takami, H. and  
 Horikoshi, K.  
 Novel oxidatively stable subtilisin-like serine proteases from  
 alkaliphilic bacillus spp.: enzymatic properties, sequences, and  
 evolutionary relationships  
 Biochem. Biophys. Res. Commun. 279 (2), 313-319 (2000)  
 20568675  
 11118284  
 2 (bases 1 to 1299)  
 Saeki, K.  
 Direct Submission  
 Submitted (20-JUL-2000) Katsuhisa Saeki, Kao corporation,  
 Biological Science Laboratory, Ichikaimachi Akabane 2606, Haga,  
 Tochigi 321-3497, Japan (E-mail:387185@kasanet.kao.co.jp,  
 Tel:81-285-68-7400, Fax:81-285-68-7403)  
 Location/Qualifiers  
 1. .1299  
 /organism="Bacillus sp. Y"  
 /mol\_type="genomic DNA"  
 /strain="Y"  
 /db\_xref="taxon:133779"  
 1. .1299  
 /gene="PROC"  
 <1. .>1299  
 /gene="PROC"  
 /codon\_start=1  
 /evidence="experimental"  
 /transl\_table=11  
 /product="protease"  
 /protein\_id="BAB21267.1"  
 /db\_xref="GI:12381941"  
 /translation="NDVARGIVKADVAQNRYGLYGGQGVAVADTGLDTRNDSSWHE  
 AFRKITALYALGRNNASDPNGHGTAVAGSVLGNALNKGAPQANLVFQSIIMDSGG  
 LGGLPSNLNTLPSQAWNAGARLHTNSWGPVNGATVANSRQVDEYVRNDWTVLFRAG

## FEATURES

source  
 gene  
 CDS

## ORIGIN

Alignment Scores:  
 Pred. No.: 5,77e-120 Length: 1299  
 Score: 2033.00 Matches: 384  
 Percent Similarity: 94.46% Conservative: 25  
 Best Local Similarity: 88.68% Mismatches: 24  
 Query Match: 90.28% Indels: 0  
 DB: 1 Gaps: 0  
 US-09-985-689A-7 (1-433) x AB046404 (1-1299)

QY 1 AsnAspValAlaArgGlyLleValIysAlaAspValAlaGlnAsnAsnPheGlyLeuTy 20  
 Db 1 AATGATGTAGCAAGAGGATAGTAAAGCTGATGTTGCACAAACAAATTCGGATTATAT 60  
 QY 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40  
 Db 61 CGACAAGGTCAAGTAGTTGCAGTAGCGACACAGGCTTAGATACAGGTCGTAAAGTAGT 120  
 QY 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyAlaLeuGlyArgThrAsn 60  
 Db 121 TCTATGTCATGAAGCAATTCGGGGAAATCACAGCTCTTACGCTTAGGAAGAACTAAT 180  
 QY 61 AsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnAla 80  
 Db 181 AATCGAGTGTATCCGAATGGGCATGGCACACATGTAGCAGGTTCTGTACTTGGTAATGCT 240  
 QY 81 ThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSerGly 100  
 Db 241 TTAATAAAGGAATGGCTCCGCAAGCTAATCTTAGCTTCCAACTATATTGGATAGCAGC 300  
 QY 101 GlyGlyLeuGlyGlyLeuProAlaAsnLeuGlnThrLeuPheSerGlnAlaTySerAla 120  
 Db 301 CGAGGATAGGTGGCTTACCATCGAATTAATACGTTATTTAGTCAAGCTTGAATGCT 360  
 QY 121 GlyAlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyThrThrAsp 140  
 Db 361 GGAGCAAGAAATTCATCTACTCTTGGGGAGCCCAAGTAATGAGGCGTACACTGCTAAC 420  
 QY 141 SerArgAsnValAspAspTyValArgLysAsnAspMetThrIleLeuPheAlaAlaGly 160  
 Db 421 TCGAGACAAAGTGAATGAATATGTTCGAAATTAATGATATGACCGTACTTTTTCAGCTGCT 480  
 QY 161 AsnGlyGlyProGlySerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThr 180  
 Db 481 AATGAAGTCTCCTAATTCAGGAACAAATAGTCTCCAGGTACAGCGAAAATGCTATTACG 540  
 QY 181 ValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyAlaAspAsnIleAsnHis 200  
 Db 541 GTGGCGCAACGGAACATATCGGCCCAAGCTTCGGTTCGATAGCAGATACCCAAATCAT 600  
 QY 201 ValAlaGlnPheSerSerArgGlyProThrArgAspGlyArgIleLysProAspValMet 220  
 Db 601 ATTGCACAATTTTCAATCGAGAGGAGCTACGAGGATGAGCAAGATTAGCCCTGACGTAA 660  
 QY 221 AlaProGlyThrTyIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTy 240  
 Db 661 GCTCCTCGGAACATTTATTTTATCAGCAGCTTCTTCTCTAGCTCCAGACTCTTCGTTTGG 720  
 QY 241 AlaAsnHisAspSerLysTyAlaTyMetGlyGlyThrSerMetAlaThrProIleVal 260  
 Db 721 GGAATATTATACAGTAATAATCGGTATATGGCGGTACCTCCATGGCGACACCTATTGTT 780  
 QY 261 AlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyValThrProLys 280  
 Db 781 GCAGGGAATGTCGCGCAATTCAGTGAGCATTTTATAAAAAATAGAGGTATTACTCTCTAAG 840

NEGPNSTISAPGTAKNAITVGTATENYRPSFGSIADNPNTAOFSSRGATRDGRIPED  
 VTACGTFTLSRSLSAPSSFWANYSKYAMGSGTSMATPIVAGNVAOLREHFKRKG  
 ITPKPSLLKALLAGATDVGCTYENGCGWGRVTINKSLNVAIYNEATALATCAQKATY  
 SFQAQAGPKIKISLWTDIAPGTSYTLVNDLDELITVAPNGQKIVGNDFSYFDNNW  
 DGRNVNENFVINAPOSQGTIIIEVQAYNVPSQFQFSLAIVH"

```

Qy 281 ProSerLeuLeuLysAlaAlaLeuLeuAlaGlyAlaAlaAspValGlyLeuGlyPhePro 300
Db 841 CCITCTTTAATAAAGCTGCACCTTATCGTGGTGTACTCATGCTGGTTTAGGATATCCT 900
Qy 301 AsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaPheVal 320
Db 901 AATGGTCACCAAGCTGGGGCGGTGTACTCTTAATAAATCGTTAAATGATAGCGTATGTC 960
Qy 321 AsnGluThrSerProLeuSerThrSerGlnLysAlaAlaThrThrSerPheThrAlaGlnAla 340
Db 961 AATCAAGCAACTGCATTAGCCACAGGACAAAAGCAACGATTTCGTTCCAGCACAAAGCG 1020
Qy 341 GlyLysProLeuLysIleSerLeuValTrpSerAspAlaProGlySerThrThrAlaSer 360
Db 1021 GGTAAACCTTTAAAAATCTCGTTAGTATGACACATGCTCCTGGAAGTAACTGATGCT 1060
Qy 361 LeuThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrLysTrpVal 380
Db 1081 TATACACTAGTTAATGATTAGATCTAGTTATTAATCTCGAATGGACAAAATATGTA 1140
Qy 381 GlyAsnAspPheThrAlaProTyAspAsnAsnTrpAspGlyArgAsnAsnValGluAsn 400
Db 1141 GGAAATGATTTTAGTTATCTCTTATGATAAATAACTGGGATGGTGCACAAATGTTGGAAC 1200
Qy 401 ValPheIleAsnAlaProGlnSerGlyThrThrValGluValGlnAlaTyAsnVal 420
Db 1201 GTATTTATAACGCTCCCAATCTGGAACTGATATTAATGGGTTCAAGGATATATGTA 1260
Qy 421 ProValSerProGlnThrPheSerIleAlaIleValHis 433
Db 1261 CCATCTGCCACACAGCGTTTCTCACTAGCTATCGTACAT 1299

RESULT 14
E03808
LOCUS      E03808          2218 bp      DNA      linear      PAT 29-SEP-1997
DEFINITION DNA encoding alkaline protease Ya.
ACCESSION  E03808
VERSION    E03808.1 GI:2172022
KEYWORDS   JP 1992197182-A/1.
SOURCE     Bacillus sp.
ORGANISM   Bacillus sp.
REFERENCE  1. (bases 1 to 2218)
AUTHORS    Tobe,S., Odera,M. and Asai,Y.
TITLE      DNA CODING ALKALINE PROTEASE YA ENZYME AND PRODUCTION OF ALKALINE
           PROTEASE YA USING THE DNA
JOURNAL    Patent: JP 1992197182-A 1 16-JUL-1992;
           LION CORP
COMMENT    OS Bacillus sp.
           PN JP 1992197182-A/1
           PD 16-JUL-1992
           PF 28-NOV-1990 JP 1990327110
           PI TOBE SEIICHI, ODERA MOTOYASU, ASAI YOSHIO
           PC C12N15/57,C11D3/386,C12N9/54,(C12N15/57,C12R1:07),(C12N9/54,
           PC C12R1:07);
           CC strandedness: Double;
           CC topology: Linear;
           CC hypothetical: No;
           CC anti-sense: No;
           CC *source: strain=J strain;
           CC *source: clone=JUB9A;
           PH Key
           PH Location/Qualifiers
           FT CDS
           FT sig_peptide
           FT mat_peptide
           FT Location/Qualifiers
           1..2218
           /organism="Bacillus sp."
           /mol_type="genomic DNA"
           /db_xref="taxon:1409"
FEATURES
source

```

```

ORIGIN
Alignment Scores: 1.29e-119 Length: 2218
Pred. No.: 2032.00 Matches: 384
Score: 94.46% Conservative: 25
Percent Similarity: 88.68% Mismatches: 24
Best Local Similarity: 90.23% Indels: 0
Query Match: 6 Gaps: 0
DB:
US-09-985-689A-7 (1-433) x E03808 (1-2218)
Qy 1 AsnAspValAlaAspGlyIleValLysAlaAspValAlaGlnAsnAsnPheGlyLeuTrp 20
Db 824 AATGATGATAGCAGAGGATAGTAAAGCTGATGTGCACAAACAATATCGGATATAT 883
Qy 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 884 GGCAAGGTCAACTAGTTGCAGTAGCAGACACAGGCTTAGATACAGGTCGTAACGATAGT 943
Qy 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyAlaLeuGlyArgThrAsn 60
Db 944 TCTATGCATGAAGCAATTCGCGGGAAATACAGCTCTTTACGCGTTAGGAGAACTAAT 1003
Qy 61 AsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnAla 80
Db 1004 AATGCGAGTGATCCGAATGGGCATGGCACACATGTAGCAGGTTCTGTACTTGGTAATGCT 1063
Qy 81 ThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSerGly 100
Db 1064 TTAAATAAAGGAATGGCTCCGCAAGCTAACTTAGTCTTCCAACTCTATTATGGATAGCAGC 1123
Qy 101 GlyGlyLeuGlyGlyLeuProAlaAsnLeuGlnThrLeuPheSerGlnAlaTySerAla 120
Db 1124 GGAGGATTAGTGGCTTACCATCGAATTAATACTGTTATTAGTCAAGCTTGAATGCT 1183
Qy 121 GlyAlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyThrThrAsp 140
Db 1184 GGAGCAAGAAATTCATACTAACTCTTGGGGAGCCCGAGTAATGGAGCGTACACTCCTAAC 1243
Qy 141 SerArgAsnValAspAspTyValArgLysAsnAspMetThrIleLeuPheAlaIleGly 160
Db 1244 TCGAGCAAGTGGATGAGTATGTTCCGAATAATATGATATGACGGTACTTTTGCAGCTGGT 1303
Qy 161 AsnGlyLysProGlySerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThr 180
Db 1304 AATGAAGGTCTCAATTCAGGAACAATTAGTCTCCAGGTACAGCGAAATGCTATTACG 1363
Qy 181 ValGlyValThrGluAsnLeuArgProSerPheGlySerTyAlaAspAsnIleAsnHis 200
Db 1364 GTCCGCGCAACGGAAACCTATCGCCACAGCTTCGGTTCGATAGCAGATAACCCAAATCAT 1423
Qy 201 ValAlaGlnPheSerSerArgGlyProThrArgAspGlyArgIleLysProAspValMet 220
Db 1424 ATTGCACAATTTTCATCGAGAGGAGCTACGAGGGATGGACGAATTAAGCCTGACGTAACA 1483
Qy 221 AlaProGlyThrTyThrIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrp 240
Db 1484 GCTCCTGGAAACATTTATTTTATCAGACCGCTTTCTTCCTAGCTCCAGACTCTTCGTTGG 1543
Qy 241 AlaAsnHisAspSerLysTyThrAlaTyMetGlyThrSerMetAlaThrProIleVal 260
Db 1544 GCGAATTTATAACAGTAAATACGCTATATATGGCGGTACCTCCATGGCGACACCTATTGTT 1603
Qy 261 AlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyValThrProLys 280
Db 1604 GCAGGGAATGTCCGCAATTAGCTGAGCATTTTATAAAAAATAGAGGTATTACTCTCTAG 1663
Qy 281 ProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspValGlyLeuGlyPhePro 300
Db 1664 CCTTCTTTAATAAAGCTGCACCTTATCGTGGTGTCTACTGATGTTGTTGGTATATCCT 1723
Qy 301 AsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaPheVal 320

```



```
Db 1724 AGTGTGACCAAGCGTGGGCGTGTACTAGATAAATCGTTAAATGAGCGTATGTC 1783
QY 321 AaGluThrSerProLeuSerThrSerGlnLysAlaThrTySerPheThrAlaGlnAla 340
Db 1784 AATGAAGCAACTGATAGCCACAGACAAAAGCAACGATTCGTTCCAGACACAGCG 1843
QY 341 GlyLysProLeuLysLysSerLeuValTrpSerAspAlaProGlySerThrAlaSer 360
Db 1844 GATAACCTTTAAATATCTGTTAGTAGACAGATGCTCTGGAAGTCAACTGCACT 1903
QY 361 LeuThrLeuValAsnAspLeuAspLeuValLleThrAlaProAsnGlyThrLysTyVal 380
Db 1904 TATACACTAGTTAATGATTTAGATCTAGTTATTACTGCTCCGATGACAAAATATGTA 1963
QY 381 GlyAsnAspPheThrAlaProTyAspAsnAsnTrpAspGlyAscAsnAsnValGluAsn 400
Db 1964 GGAATGATATTAGTTATCTTATGATAATAACTGGGATGTCGCAACAATGTTGGAAC 2023
QY 401 ValPheIleAsnAlaProGlnSerGlyThrTyThrValGluValGlnAlaTyAsnVal 420
Db 2024 GTATTATAAAGCTCCGCAATCTGGAACGTATATAATGAGGTTCAAGCGTATAATGTA 2083
QY 421 ProValSerProGlnThrPheSerLeuLalleValHis 433
Db 2084 CCATCTGCCACACAGCGTTTCTCCTAGTATCTGATACAT 2122

RESULT 15
AF268611/c 60006 bp DNA linear BCT 30-MAY-2001
LOCUS
DEFINITION
Uncultured marine group II euryarchaeote EBAC37F11, BAC clone
37F11, partial sequence.
ACCESSION
AF268611
VERSION
GI:9664575
KEYWORDS
uncultured marine group II euryarchaeote 37F11
SOURCE
Archaea; Euryarchaeota; Marine Group II; environmental samples.
ORGANISM
1 (bases 1 to 60006)
Beja, O., Suzuki, M.T., Koonin, E.V., Aravind, L., Hadd, A.,
Nguyen, L.P., Villacorta, R., Amjadi, M., Garrigues, C.,
Jovanovich, S.B., Feldman, R.A. and Delong, E.F.
Construction and analysis of bacterial artificial chromosome
libraries from a marine microbial assemblage
Environ. Microbiol. 2 (5), 516-529 (2000)
21128653
11233180
2 (bases 1 to 60006)
Beja, O., Suzuki, M.T., Koonin, E.V., Aravind, L., Hadd, A.,
Nguyen, L.P., Villacorta, R., Amjadi, M., Garrigues, C.,
Jovanovich, S.B., Feldman, R.A. and Delong, E.F.
Direct Submission
Submitted (17-MAY-2000) R & D, Monterey Bay Aquarium Research
Institute, P.O. Box 628, Moss Landing, CA 95039-0628, USA
FEATURES
location/Qualifiers
1..60006
/organism="uncultured marine group II euryarchaeote 37F11"
/mol_type="genomic DNA"
/db_xref="taxon:133822"
/cldone="BAC 37F11"
/note="environmental sample"
1484..3178
/note="37F11#1"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="S-layer protein"
/protein_id="AAF97179.1"
/db_xref="GI:9664576"
/translacion="MEPQCTELLIDIFVQSPSSAIAQVNIIVKRVSDGSCQGLEBEI
FVRVSPFENLQASQWFWGGGYPLANNENTGNTMWSHTITLNLPGWNISGMQ
SASVSPHARGIPPEIIPPTNWDKQPLELSIVHPQGLVMYVESITVHNSIFASTP
VLGSGAGSQAFSLHLSINPSTVTVPATMNQDEYELILEEGINFETISSTNDIAIQ
```

CDS

```
/note="37F11#2"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="unknown"
/protein_id="AAF97180.1"
/db_xref="GI:9664577"
/translacion="MNQSNSTSAFYTLDCVSENLEDTFOALLEIRGNHQLWEKLV
NREAFTRKKQGLDIDDRKNTWDDATLWFCVRLNDRDSAPITLNGENKRRFGR
FPHTGSLQILENLSIPQSEGLYDLIEHLYLTERCSMTNGTKFLEKAGL
MILGYITFEVKALRSMLSGMWVSEEPDGGVREAVRHLLNALLMAAERRGAGLIH
RQHR"
complement (3892..4797)
/note="37F11#3"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="SAICAR synthase"
/protein_id="AAF97181.1"
/db_xref="GI:9664578"
/translacion="MGEMLYQGVKQVWSTDDPILFRFTNOISVFDQILPSLPK
LEFNRTYTHWFKLVEQGVGTHLVELNAPDRCLVKVEVVEPGAIPEDAEWVFP
GESNRYLTSLSAWRFORGELTAESLQSDAYGVKLESEFVETTPKEADRID
REALQISNTIDQFSDISVQAALKYDEIIEERAAKGLHVDGKKGKFGKGVLV
DPTGDEDRWDAAEYNGECIELSKFVTRTHYINSQHQLKLAIDSGAIDPPIPA
LPESVIBETASLYSSMYERLTSQF"
complement (4846..5508)
/note="37F11#4"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="translation initiation factor 6"
/protein_id="AAF97182.1"
/db_xref="GI:9664579"
/translacion="MPIDTLDHSGDQIGHLASVGNVLFPHPRELPTPTIEKLEQVVG
LELAPIGTSGNLGALLAGNTKGMADVATESDIDALTAYGDVWVGGVNTAGNL
LVNEQGVASPSIPKQGLDILSEVLVDIISTIGQDVGSLGVNSQGVLLHPDV
APBEVLLEBILGVPMVGVTVSGSPVVGAGICASNAGALAGSETTGPELNRIEDALG
LI"
5597..5745
/product="5S ribosomal RNA"
5816..8611
/product="23S ribosomal RNA"
complement (8888..10192)
/note="37F11#7: contains Zn-ribbon and TPR-like repeats"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="unknown"
/protein_id="AAF97183.1"
/db_xref="GI:9664580"
/translacion="MNECAVCGILYLAGSPCAGCSQOLRAQDSFSDSAAALPTEVPG
LDDAAQWYDLLEGIEAPVEQEEAASLPGFGESQTNISRLPGIGSGEGIPPEL
SNODGQSESEVVOHVEPVIESTNLEPVVALETQDPSVVSASVIEPVLAPEAPD
LPLIEEPVQVPEGPRILTAIPFBEVDTSLPNEWSIESFGSDQIFTAEDVIE
VVFSDLEDTPVHVEHGDENPQMEKQSVISERKSLQPDHLHPAQLMVDVGTDSRAL
LESQGTIGQDWRSAARSQQIVARFPQDSGANNYNIGALLQVAVSMQNSDFIYS
NSATQFEAILLSREAVTNSQPESTYNLSQALLSGREERKALLDWMAGNSQRES
NFTNLRAAILAQGRYDEAKRLVPLQADKLASGNLILKPLAL"
complement (10248..11969)
/note="37F11#8"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="permease"
/protein_id="AAF97184.1"
```

rRNA

rRNA

CDS

CDS



```
QY 257 ThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGly 276
Db 16658 ACAGCAGTAGCAGGCTGTCACATCCCTCGCAGCTGAATATCTGCGTGAAGTTGCTGGA 16599
QY 277 ValThr---ProLysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspVal 295
Db 16598 ATCAACAAACCATCTGCAAGCTCATTAAGGCAACTCTCATCAACGGGGCAGAAGATTTA 16539
QY 296 GlyLeu---GlyPheProAsnGlyAsnGlnGlyTpgLysArgValThrLeuAspLysSer 314
Db 16538 GGCACACCGGACATTCTTAACCAACAGGAGATGGGGGCAAAATTGATCTCGAAAATAGC 16479
QY 315 LeuAsnValAlaA-----PheValAsnGluThrSerProLeu 326
Db 16478 CTGAATCCTTCGCTGCGGCTTCTCTGCGATGCTTTTCAGGATGATGAGCGTGAACCTT 16419
QY 327 SerThrSerGlnLysAlaThrTyrSerPheThrAlaGlnAlaGlyLysProLeuLysIle 346
Db 16418 CAGGCGAGGATTCTCATTTATTTCTTTCACCTTGATGATCAAGGGTATTGACATT 16359
QY 347 SerLeuValTyrSerAspAlaProGlySerThrThrAlaSerLeuThr-----Leu 363
Db 16358 ACCTTGGCATGGAGCGATGCAGAGCCAGTGCATAATGCTGCACAATCTGAATCACGCTTA 16299
QY 364 ValAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrLysTyrValGlyAsnAsp 383
Db 16298 TTGAACAATTATAGATTGATTCTAATTGACCCGATGGCTCTTCTACTTGGGTAATGAC 16239
QY 384 PheThrAlaProTyrAsp-----AsnAsnTyrAspGlyA-gAsnAsnValGluAsn 400
Db 16238 TTCTCTTCTGGGATTTCACACACAGCGGTTCCGCGTGATAATCTGAACAATATCGAGCGC 16179
QY 401 ValPheIleAsnAla-----ProGlnSerGlyThrTyrThrValGluValGlnAla 417
Db 16178 ATTCGCATCCCTGCAGGAGCGACGACACAGAACCGTGATTGGATCGTTACAGTTGAA--- 16122
QY 418 TyrAsnValProValSerProGlnThrPheSerLeuAlaIle 431
Db 16121 ---CATCGAGGAGGAAGTTCTCAGCGCTACAGTATCGTTATT 16083
```

Search completed: April 4, 2004, 08:10:14  
Job time : 3595.27 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: April 3, 2004, 23:50:44 ; Search time 63.4512 Seconds  
(without alignments)  
3787.066 Million cell updates/sec

Title: US-09-985-689A-7  
Perfect score: 2252  
Sequence: 1 NDVARGIVKADVAQNFGILY.....EVQAYNPVSPQTFSLAIVH 433

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame\_plus\_p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO.spool/US09985689/runat\_31032004.161807.4184/app\_query.fasta\_1.3498  
-DB=Issued Patents NA -Qfmt=fastap -SUFFIX=rni -MNMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09985689 @C@N 1.1 142 @runat\_31032004.161807.4184 -NCPU=6 -ICPU=3  
-NO.MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents NA:  
1: /cgn2\_6/ptodata/2/ina/5A COMB seq.\*  
2: /cgn2\_6/ptodata/2/ina/5B COMB seq.\*  
3: /cgn2\_6/ptodata/2/ina/6A COMB seq.\*  
4: /cgn2\_6/ptodata/2/ina/6B COMB seq.\*  
5: /cgn2\_6/ptodata/2/ina/PTUS COMB seq.\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2252	100.0	3003	2	US-08-873-479-41 Sequence 41, Appl
2	2125.5	94.4	1923	4	US-09-509-814A-5 Sequence 5, Appl
3	2120.5	94.2	1923	4	US-09-509-814A-7 Sequence 7, Appl
4	2110.5	93.7	1920	4	US-09-509-814A-3 Sequence 3, Appl
5	432.5	19.2	1977	3	US-08-894-818B-2 Sequence 2, Appl
6	432.5	19.2	1977	4	US-09-445-472-11 Sequence 11, Appl
7	396	17.6	1236	4	US-08-445-472-2 Sequence 2, Appl
8	396	17.6	1566	3	US-08-894-818B-4 Sequence 4, Appl
9	396	17.6	1962	3	US-08-894-818B-34 Sequence 34, Appl
10	396	17.6	1962	4	US-09-445-472-15 Sequence 15, Appl
11	390	17.3	1977	3	US-08-894-818B-6 Sequence 6, Appl
12	332.5	14.8	2539	3	US-09-000-016-3 Sequence 3, Appl

13	332.5	14.8	2539	4	US-09-514-340-3 Sequence 3, Appl
14	327.5	14.5	2809	3	US-09-000-016-1 Sequence 1, Appl
15	327.5	14.5	2809	4	US-09-514-340-1 Sequence 1, Appl
16	311.5	13.8	2835	1	US-08-750-532-2 Sequence 2, Appl
17	311.5	13.8	4765	3	US-08-750-532-8 Sequence 8, Appl
18	311.5	13.8	4765	3	US-08-894-818B-7 Sequence 7, Appl
19	311.5	13.8	4765	4	US-09-445-472-5 Sequence 5, Appl
20	309	13.7	2532	1	US-07-671-376C-4 Sequence 4, Appl
21	297	13.2	1859	3	US-08-894-818B-15 Sequence 15, Appl
22	276	12.3	898	1	US-08-750-532-7 Sequence 7, Appl
23	271.5	12.1	1306	4	US-09-966-921A-1 Sequence 1, Appl
24	271.5	12.1	1330	4	US-09-966-921A-5 Sequence 5, Appl
25	270	12.0	564	1	US-08-750-532-11 Sequence 11, Appl
26	270	12.0	564	3	US-08-894-818B-14 Sequence 14, Appl
27	257.5	11.4	807	1	US-07-706-691G-56 Sequence 56, Appl
28	257.5	11.4	807	1	US-08-254-021-56 Sequence 56, Appl
29	257.5	11.4	807	3	US-08-618-446-56 Sequence 56, Appl
30	257.5	11.4	807	3	US-08-980-135-56 Sequence 56, Appl
31	257.5	11.4	807	4	US-09-585-798-56 Sequence 56, Appl
32	256.5	11.4	807	1	US-08-566-369-2 Sequence 2, Appl
33	256.5	11.4	807	3	US-09-074-331-2 Sequence 2, Appl
34	256.5	11.4	807	5	FCT-US95-01937-2 Sequence 2, Appl
35	255.5	11.3	807	1	US-07-706-691G-57 Sequence 57, Appl
36	255.5	11.3	807	1	US-08-254-021-57 Sequence 57, Appl
37	255.5	11.3	807	2	US-08-618-446-57 Sequence 57, Appl
38	255.5	11.3	807	3	US-08-980-135-57 Sequence 57, Appl
39	255.5	11.3	807	4	US-09-585-798-57 Sequence 57, Appl
40	253.5	11.3	807	1	US-08-566-369-5 Sequence 5, Appl
41	253.5	11.3	807	3	FCT-US95-01937-5 Sequence 5, Appl
42	253.5	11.3	807	5	FCT-US95-01937-5 Sequence 5, Appl
43	253	11.2	10216	2	US-08-875-154-1 Sequence 1, Appl
44	251.5	11.2	807	1	US-08-566-369-8 Sequence 8, Appl
45	251.5	11.2	807	3	US-09-074-331-8 Sequence 8, Appl

ALIGNMENTS

RESULT 1  
US-08-873-479-41  
; Sequence 41, Application US/08873479  
; Patent No. 5891701  
; GENERAL INFORMATION:  
; APPLICANT: Sloma, Alan  
; APPLICANT: Lyne, Christian  
; TITLE OF INVENTION: Nucleic Acids Encoding A Polypeptide  
; TITLE OF INVENTION: Having Protease Activity  
; NUMBER OF SEQUENCES: 57  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: No. 58917010 No. 5891701disk of No. 5891701th America  
; STREET: 405 Lexington Avenue  
; CITY: New York  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 10174  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/873,479  
; FILING DATE: 12-JUN-1997  
; CLASSIFICATION: 530  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Agis, Cheryl H  
; REGISTRATION NUMBER: 34,086  
; REFERENCE/DOCKET NUMBER: 5251.000-US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-857-0123  
; TELEFAX: 212-878-9655  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 41:  
; SEQUENCE CHARACTERISTICS:

41699 102b

```

; LENGTH: 3003 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
US-08-873-479-41

Alignment Scores:
Pred. No.: 1,86e-222 Length: 3003
Score: 232.00 Matches: 433
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 2 Gaps: 0

US-09-985-689A-7 (1-433) x US-08-873-479-41 (1-3003)

QY 1 AsnAspValAlaArgGlyIleValLysAlaAspValAlaGlnAsnAsnPhedGlyLeuTyr 20
Db 1470 AATGACGTGGCCGTCGTCATTGTGAAGACAGACGTCGCACAAATAACTTTGGCTTATAT 1529

QY 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 1530 GGACAGGACAGATGTAGCAGTGTCTGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1589

QY 41 SerMetHisGluAlaPheArgGlyIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
Db 1590 TCGATGTCATGAAGCATTCCGCGGTAAAGATTACCGCACTATATGCTGGCGCAGACGAAT 1649

QY 61 AsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyValAla 80
Db 1650 AACGCCAATGATCCAAATGGACATGGACATGGACATGGACATGGACATGGACATGGACATGG 1709

QY 81 ThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSerGly 100
Db 1710 ACAATTAAGGATGGACCGCACCGCAATCTAGTCTTCAATCTATATGATAGTAGTGGT 1769

QY 101 GlyGlyLeuGlyGlyLeuProAlaAsnLeuGlnThrLeuPheSerGlnAlaTyrSerAla 120
Db 1770 GGAGGGCTGGGAGGACTACCTGCTAATCTCAAAATTAATTCAGTCAAGCAATATAGTCT 1829

QY 121 GlyAlaArgIleHisThrAsnSerTyrGlyAlaProValAsnGlyAlaTyrThrAsp 140
Db 1830 GGAGCGCAATTCATCAAGATTCATGGGGGCTCCAGTAAACGGTGCCTATACGACAGAC 1889

QY 141 SerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaGly 160
Db 1890 TCTCGAATGTGTGATGATTATGTGAGAAAAATGATATGACGATCTCTTTTGGCGCGGA 1949

QY 161 AsnGluGlyProGlySerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThr 180
Db 1950 AATGAGGACACAGGAGTAGCGGTACATCAATCAATGTCACAGGACAGCAAAAAATCGCAT 2009

QY 181 ValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsnHis 200
Db 2010 GTTGGGGCAACGAAAAACCTACGTCAGCTTCGGATCTTATGCGGATAAATATTAACCAT 2069

QY 201 ValAlaGlnPheSerSerArgGlyProThrArgAspGlyValGlyLeuLysProAspValMet 220
Db 2070 GTTGCTCAATCTCTTCACAGAGTCTCTACTAGATGAGCGTATTAAACCGGACGCTCATG 2129

QY 221 AlaProGlyThrTyrIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTyr 240
Db 2130 GCACCAAGGTACGTATATCTCTCTGCTAGATCATCATTAGCTCCAGATTCCTCTCTGG 2189

QY 241 AlaAsnHisAspSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProIleVal 260
Db 2190 GCAACCATCATGATAGTAAATATGCTACATGGGTGGTACTCTTATGCTACTCTCAATGTA 2249

QY 261 AlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyValThrProLys 280
Db 2250 GCAGGTAAATGTTCCACAAATTAAGGGAGCATTTTGTGAAAAATAGAGGGGTAACTCTTAAG 2309
```

```

281 ProSerLeuLeuLysAlaAlaLeuLysAlaGlyAlaAlaAspValGlyLeuGlyPhePro 300
2310 CCTTCCTTTTAAAGCTGCTTAATTCAGAGTGTGCGGATGTGGACTTGGCTTTTCCA 2369

QY 301 AsnGlyAsnGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsnValAlaPheVal 320
2370 AATGTAACCAAGATGGGAGAGTAACTGTAGTAAATCCCTTAATGTCGCAATTTGTG 2429

QY 321 AsnGluThrSerProLeuSerThrSerGlnLysAlaThrTyrSerPheThrAlaGlnAla 340
2430 AATGAAACGAGCCCTTTATCAACAAAGTCAAAAAGCAACATATTCGTTACGGCTCAAGCT 2489

QY 341 GlyLysProLeuLysIleSerLeuValTyrSerAspAlaProGlySerThrThrAlaSer 360
2490 GGTAACCCCTTAAATAATATCATCTTTTGTGATGATGATGATGATGATGATGATGATGATG 2549

QY 361 LeuThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrLysTyrVal 380
2550 CTAACTTTAGTGAATGATTTAGACTTATCACTGACCAATGGAACCTTAATATCGCTC 2609

QY 381 GlyAsnAspPheThrAlaProTyrAspAsnAsnTrpAspGlyArgAsnAsnValGluAsn 400
2610 GGAATGACTTTACAGCACCGTATGATTAACAAATTTGGGATGGCAGAAACACAGCTGGAAT 2669

QY 401 ValPheIleAsnAlaProGlnSerGlyThrTyrThrValGluValGlnAlaTyrAsnVal 420
2670 GTGTTTATCAATGCTCTCTCAAGCGGAACGTATACAGTCGAAGTGCAGGCTTACAATGTA 2729

QY 421 ProValSerProGlnThrPheSerLeuAlaIleValHis 433
2730 CCAGTAAGTCCGCAAAACCTTTTCTTTAGCGATTTGTACAT 2768

RESULT 2
US-09-509-814A-5
; Sequence 5, Application US/09509814A
; Patent No. 6376227
; GENERAL INFORMATION:
; APPLICANT: TAKAIWA, MIKIO
; APPLICANT: OKUDA, MITSUYOSHI
; APPLICANT: SAEKI, KATSUHIKA
; APPLICANT: KUBOTA, HIROMI
; APPLICANT: HITOMI, JUN
; APPLICANT: KAGEYAMA, YASUSHI
; APPLICANT: SHIKATA, SHITSUW
; APPLICANT: NOMURA, MASAFUMI
; TITLE OF INVENTION: ALKALINE PROTEASE
; FILE REFERENCE: 0327-0832-OPCT
; CURRENT APPLICATION NUMBER: US/09/509,814A
; CURRENT FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: PCT/JP98/04528
; PRIOR FILING DATE: 1998-10-07
; PRIOR APPLICATION NUMBER: JP 9-274570
; PRIOR FILING DATE: 1997-06-08
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 5
; LENGTH: 1923
; TYPE: DNA
; ORGANISM: Bacillus sp.
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1923)
US-09-509-814A-5

Alignment Scores:
Pred. No.: 1.07e-209 Length: 1923
Score: 2125.50 Matches: 406
Percent Similarity: 97.93% Conservatives: 19
Best Local Similarity: 97.55% Mismatches: 8
Query Match: 94.38% Indels: 1
DB: 4 Gaps: 1

US-09-985-689A-7 (1-433) x US-09-509-814A-5 (1-1923)
```

QY 1 AsnAspValAlaArgGlyIleValIleValAlaAspValAlaGlnAsnAsnPheGlyLeuTyr 20  
Db 619 AATGATGTGGCGGTGGAAATGTGCAAGCGGATGTGGCTCAGACGAGCTACGGGTGTAT 678  
QY 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40  
Db 679 GGACAAAGACAGATCGTAGTGGGTGGCGATACAGGGCTTGATACAGGTGCGCAATGACAGT 738  
QY 41 SerMetHisGluAlaPheAArgGlyIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60  
Db 739 TCGATGATGATGAGCGCTTCCGCGGAAATTTACTGCAATATATGCAATGGGACGAGCAAT 798  
QY 61 AsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn--- 79  
Db 799 AATGCCAATGATACGAATGTGTATGTACGATGTGGCTGGCTTCCGTAATGGAACCGC 858  
QY 80 AlaThrAsnIysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 99  
Db 859 TCCACTAATAAGGAATGGCGCTCAGCGGAATCTAGTCTTCAATCTATCATGATGATAGC 918  
QY 100 GlyGlyGlyLeuGlyGlyLeuProAlaAsnLeuGlnThrLeuPheSerGlnAlaTyrSer 119  
Db 919 GGTGGGAGCTTGAGAGACTACTTGTGAATCTGCAACCTTATTCAGCCAGCATACAGT 978  
QY 120 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyrThr 139  
Db 979 GCTGGTCCGAATTCATACAAACTCTGGGGAGCAGCAGTGAATGGGCTTTACACAACA 1038  
QY 140 AspSerArgAsnValAspAspTyrValArgIysAsnAspMetThrIleLeuPheAlaAla 159  
Db 1039 GATTCAGGAATGTGGATGACTATGTGGCAAAATGATATGACCATCTTTTCCGTGCC 1098  
QY 160 GlyAsnGlyProGlySerGlyThrIleSerAlaProGlyThrAlaIysAsnAlaIle 179  
Db 1099 GGGAAATGAAGACCGAAGCGCGAACCATCAGTGCAACGAGGACAGCTAAAATGCAATA 1158  
QY 180 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 199  
Db 1159 ACAGTCGAGCTACGGAACCTCCGCCAAGCTTGGTCTTATGCGGACATATCAAC 1218  
QY 200 HisValAlaGlnPheSerSerArgGlyProThrArgAspGlyArgIleIysProAspVal 219  
Db 1219 CATGTGCGACAGTTCTCTTCCAGTGGACCGGACAAAGATGCGAGCATCAACCGGATGTC 1278  
QY 220 MetAlaProGlyThrTyrIleLeuSerAlaArgSerSerLeuAlaProAspSerPhe 239  
Db 1279 ATGSCACCGGAAAGTTTCACTATCATGCAAGATCTTCTTTCGACCGGATTCCTCTTC 1338  
QY 240 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIle 259  
Db 1339 TGGCGAACCATGACAGTAATAATGATACATGGGTGGAACTCCATGCTGCTACACCGATC 1398  
QY 260 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValIysAsnArgGlyValThrPro 279  
Db 1399 GTTGTGAAACGTGGGCACAGCTTCGTGAGCATTTTGTGAAAACACAGAGGCATCACCA 1458  
QY 280 LysProSerLeuLeuIysAlaAlaLeuIleAlaGlyAlaAlaAspValGlyLeuGlyPhe 299  
Db 1459 AAGCTTCTCTATTAAGCGGACACTGATGCGGTCGAGCGTACATCGGCTTGGCTTAC 1518  
QY 300 ProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaPhe 319  
Db 1519 CCGAACCGTAACCAAGATGGGACGAGTGATGATGATAAATCCCTGAAGCTTGGCTTAT 1578  
QY 320 ValAsnGluThrSerProLeuSerThrSerGlnIysAlaThrTyrSerPheThrAlaGln 339  
Db 1579 GTGAAACGAGTCCAGTCTCTATCCACGACCCAAAAGCGACGATCTCGTTTACTGCTACT 1638  
QY 340 AlaGlyLysProLeuLysIleSerLeuValTrpSerAspAlaProGlySerThrAla 359  
Db 1639 GCGGCAAGCCTTGAATACTCCCTGGTATGCTGTGATGCCCTCGGAGCAGCAACTGCT 1698

QY 360 SerLeuThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrLysTyr 379  
Db 1699 TCCGTAAACGCTTGTCAATGATCGACCTTGTCAATACCGCTCCAAATGGCACAGTAT 1758  
QY 380 ValGlyAsnAspPheThrAlaProTyrAspAsnAsnTrpAspGlyArgAsnValGlu 399  
Db 1759 GTAGGAATGACTTTTACTTCGCCATACATGTAACCTGGGATGGCCCAATAACGTAGAA 1818  
QY 400 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrValGluValGlnAlaTyrAsn 419  
Db 1819 AATGATATTATTAAATGCACCAAAAGCGGACGATATACAATTGAGGTACAGGCTTAAC 1878  
QY 420 ValProValSerProGlnThrPheSerLeuAlaIleValHis 433  
Db 1879 GTACCGTTGGACACAGACCTTCTCGTTGCCAATTGTGAAT 1920  
RESULT 3  
US-09-509-814A-7  
; Sequence 7, Application US/09509814A  
; Patent No. 6376227  
; GENERAL INFORMATION:  
; APPLICANT: TAKAIWA, MIKIO  
; APPLICANT: OKUDA, MITSUYOSHI  
; APPLICANT: SAEKI, KATSURISHA  
; APPLICANT: KOBOTA, HIROMI  
; APPLICANT: HITOMI, JUN  
; APPLICANT: KAGEYAMA, YASUSHI  
; APPLICANT: SHIKATA, SHITSUMI  
; APPLICANT: NOMURA, MASAFUMI  
; TITLE OF INVENTION: ALKALINE PROTEASE  
; FILE REFERENCE: 0327-0832-OPCT  
; CURRENT APPLICATION NUMBER: US/09/509,814A  
; CURRENT FILING DATE: 2000-04-06  
; PRIOR APPLICATION NUMBER: PCT/JP98/04528  
; PRIOR FILING DATE: 1998-10-07  
; PRIOR APPLICATION NUMBER: JP 9-274570  
; PRIOR FILING DATE: 1997-06-08  
; NUMBER OF SEQ ID NOS: 24  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 7  
; LENGTH: 1923  
; TYPE: DNA  
; ORGANISM: Bacillus sp.  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(1923)  
US-09-509-814A-7  
Alignment Scores:  
Pred. No.: 3,528-209 Length: 1923  
Score: 2120.50 Matches: 405  
Percent Similarity: 97.70% Conservative: 19  
Best Local Similarity: 93.32% Mismatches: 9  
Query Match: 94.16% Indels: 1  
DB: 4 Gaps: 1  
US-09-985-689A-7 (1-433) x US-09-509-814A-7 (1-1923)  
QY 1 AsnAspValAlaArgGlyIleValIysAlaAspValAlaGlnAsnAsnPheGlyLeuTyr 20  
Db 619 AATGATGTGGCGGTGGAAATGTGCAAGCGGATGTGGCTCAGACGAGCTACGGGTGTAT 678  
QY 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40  
Db 679 GGACAAAGACAGATCGTAGTGGGTGGCGATACAGGGCTTGATACAGGTGCGCAATGACAGT 738  
QY 41 SerMetHisGluAlaPheAArgGlyIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60  
Db 739 TCGATGATGATGAGCGCTTCCGCGGAAATTTACTGCAATATATGCAATGGGACGAGCAAT 798  
QY 61 AsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn--- 79  
Db 799 AATGCCAATGATACGAATGTGTATGTACGATGTGGCTGGCTTCCGTAATGGAACCGC 858

```
QY 80 AlaThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 99
Db 859 TCACATATATAGGAATGGCGCTCAGCGCAATCTAGTCTTCCAACTCATGCGATAGC 918
QY 100 GlyGlyGlyLeuGlyLeuProAlaAsnLeuGlnThrLeuPheSerGlnAlaTyrSer 119
Db 919 GGTGGGGGACTTGGAGGACTACCTTCGAATCTCGAAACCTTATTCAGCCAGCATACAGT 978
QY 120 AlaGlyAlaArgIleHisThrAsnSerTrpGlyValAlaProValAsnGlyAlaTyrThr 139
Db 979 CTGGTCCAGAAATCTATCAAACTCTGGGAGCAGCAGTGAATGGGGCTTACACACA 1038
QY 140 AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 159
Db 1039 GATTCCGAAATGTGGATGACTATGTGCGCAAAATGATATGACGATCTCTTTTCGTGCG 1098
QY 160 GlyAsnGluGlyProGlySerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 179
Db 1099 GGGATGAGAGCAGCAGCGGCGGACCATCAGTCCACAGCAGCAGTAAATGCAATA 1158
QY 180 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 199
Db 1159 ACAGTCGGAGCTACGGAAACCTCGGCCAAGCTTTGGGTCTTATGCGGACAATATCAAC 1218
QY 200 HisValAlaGlnPheSerSerArgGlyProThrArgAspGlyArgIleLysProAspVal 219
Db 1219 CATGTGGCACAGTCTCTTCACGTGGACCGCAAGGATGACGATCAACCGGATGTC 1278
QY 220 MetAlaProGlyThrTyrIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 239
Db 1279 ATGGCACCGGGAAGCTTCATCTACTATCAGCAAGATCTTCTCTTGCACCGGATCTCTC 1338
QY 240 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProIle 259
Db 1339 TGGCGCAACCATGACAGTAATATGCAATACATGGGTGGAGCGTCCATGGCTTACCCGATC 1398
QY 260 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyValThrPro 279
Db 1399 GTTGTGGAAAGCTGGCACAGCTCGTGAGCATTTTGTGAAAAACAGAGGCAATCACACCA 1458
QY 280 LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspValGlyLeuGlyPhe 299
Db 1459 AAGCCTTCTCTATTAAGAGCGCACTGATTCGGGTGGAGTGCATCGCGCTTGGGTAC 1518
QY 300 ProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaPhe 319
Db 1519 CCGAACGGTAACCAAGATGGGACGAGTGCATTTGGATAATCCCTGAAAGTTCGCTAT 1578
QY 320 ValAsnGluThrSerProLeuSerThrSerGlnLysAlaThrTyrSerPheThrAlaGln 339
Db 1579 GTGAACGAGTCCAGTCTCTATCCACCAAGCAAGGAGCGATCTCTCTTACTGCTACT 1638
QY 340 AlaGlyLysProLeuLysIleSerLeuValTrpSerAspAlaProGlySerThrAla 359
Db 1639 GCCGGCAAGCCTTTGAAATACTCTCTGGTATGGTCTGATGCCCTCGGAGCAACACTGCT 1698
QY 360 SerLeuThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrLysTyr 379
Db 1699 TCCGTAAACGCTGTCAATGATCTGACCTTGTCAATACCGCTTCCAAATGGCAGACAGTAT 1758
QY 380 ValGlyAsnAspPheThrAlaProTyrAspAsnAsnTrpAspGlyArgAsnValGlu 399
Db 1759 GTAGGAATAGACTTCTTCCGCAATACATGATACTGGGATGGCGGCAATTAACGTAGA 1818
QY 400 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrValGluValGlnAlaTyrAsn 419
Db 1819 AATGTATTATTATATGCCACCAAGCGGACGATATACAAATGAAGTACAGGCTTATAAC 1878
QY 420 ValProValSerProGlnThrPheSerLeuAlaIleValHis 433
Db 1879 GTACCGGTTGGACCAAGAACTCTCGTTGGCAATTTGTGAAT 1920
```

## RESULT 4

```
US-09-509-814A-3
; Sequence 3, Application US/09509814A
; Patent No. 6376227
; GENERAL INFORMATION:
; APPLICANT: TAKAIWA, MIKIO
; APPLICANT: OKUDA, MITSUYOSHI
; APPLICANT: SAEKI, KATSUHIISA
; APPLICANT: KUBOTA, HIROMI
; APPLICANT: HITOMI, JUN
; APPLICANT: KAGEYAMA, YASUSHI
; APPLICANT: SHIKATA, SHITSUW
; APPLICANT: NOMURA, MASAFUMI
; TITLE OF INVENTION: ALKALINE PROTEASE
; FILE REFERENCE: 0327-0832-OPCT
; CURRENT APPLICATION NUMBER: US/09/509,814A
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: PCT/JP98/04528
; PRIOR FILING DATE: 1998-10-07
; PRIOR APPLICATION NUMBER: JP 9-274570
; PRIOR FILING DATE: 1997-06-08
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 3
; LENGTH: 1920
; TYPE: DNA
; ORGANISM: Bacillus sp.
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1920)
US-09-509-814A-3
```

```
Alignment Scores:
Pred. No.: 3,79e-208 Length: 1920
Score: 2110.50 Matches: 404
Percent Similarity: 97.24% Conservative: 18
Best Local Similarity: 93.03% Mismatches: 11
Query Match: 93.72% Indels: 1
DB: Gaps: 1
```

US-09-985-689A-7 (1-433) x US-09-509-814A-3 (1-1920)

```
QY 1 AsnAspValAlaArgGlyIleValLysAlaAspValAlaGlnAsnAsnPheGlyLeuTyr 20
Db 616 AATGATGTGGCAGAGGATTTGTCAAGCGGATGTGGACAGAGGAGCTACGCTTGTAT 675
QY 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 676 GGAACAGGCCAGATTCTGCAGTTGCCGATCTGGATTGGATACAGGAAGAACGACAGT 735
QY 41 SerMetHisGlnAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
Db 736 TCGATGCATGAAGCCCTCCCGGTAAATTAACAGCACTATATGCACTGGGTGGAGCAAT 795
QY 61 AsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn--- 79
Db 796 AATGCGAATGATACGAACGGTCTATGCCATGTGGCAGGTTCGGTATTAGGAATGCG 855
QY 80 AlaThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 99
Db 856 GCAACGAATAAAGAAATGGCACCTCAAGCGAATCTGGTTTTTCAATCCATCATGGATAGC 915
QY 100 GlyGlyGlyLeuGlyGlyLeuProAlaAsnLeuGlnThrLeuPheSerGlnAlaTyrSer 119
Db 916 AGTGGTGGGCTTGGAGGCTTGCCTTCCAATCTGCAAAACCTTATTAGCAAGCATTCAGT 975
QY 120 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyrThr 139
Db 976 GCAAGGTGCCAGAAATTCATCAAACTCTCTGGGGGGCAGCGGTGAATGGGGCTTACAGACA 1035
QY 140 AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 159
Db 1036 GATTCCAGAAATGTGGATGACTATGTAAGGAATAATGATATGACGATCTCTTTCGGGCT 1095
```



160 GlyAsnGluGlyProGlySerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 179  
Db 1096 GGGATGAAGAGCGGACCGGTCACCATCATGCTGACCTGTACGGCTAAAGCCGATA 1155  
Qy 180 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerThrAlaAspAsnIleAsn 199  
Db 1156 ACAGTCGGCGCAACCGCAAACTCGCTCAAGCTTCGGTCTCTATGTCAGATTAATTAAC 1215  
Qy 200 HisValAlaGlnPheSerSerArgGlyProThrArgAspGlyArgIleLysProAspVal 219  
Db 1216 CAGCTTGACAGTTCCTTCCTCCGCGCCGACAAAGATGGCGAATCAAGCCTGATGC 1275  
Qy 220 MetAlaProGlyThrThrIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 239  
Db 1276 ATGCGCCCGAGGACATACATATTTATCAGCAAGATCTTCTGTGACCGGATTCCTCTTC 1335  
Qy 240 TrpAlaAsnHisAspSerLysThrValAlaThrMetGlyGlyThrSerMetAlaThrProIle 259  
Db 1336 TGGCGCAATCATGACAGCAATATGCTATATGGGTGGACGCTCCATGGCAACCGGAT 1395  
Qy 260 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyValThrPro 279  
Db 1396 GTTCCCGGGAATTTCCAGAGCTCCGTCGAGCATTTGTGAAAAATAGAGGAATCACTCT 1455  
Qy 280 LysProSerLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspValGlyLysPhe 299  
Db 1456 AAGCTTCCTATTGAAGCAGCTTGTATGTCAGCTGCTGATGTTGGATTCGGTAT 1515  
Qy 300 ProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaPhe 319  
Db 1516 CCGAACGGAAACCAAGATGGGCGGAGTGAGCCCTGGATAATCGTTGAAGTTGGCTAT 1575  
Qy 320 ValAsnGlyThrSerProLeuSerThrSerGlnLysAlaThrThrSerPheThrAlaGln 339  
Db 1576 GTGAACGAATCCAGTCCCTATCAACTAGCAAAAGGACATATACCTTTACTGCAACG 1635  
Qy 340 AlaGlyLysProLeuLysIleSerLeuValThrSerAspAlaProGlySerThrThrAla 359  
Db 1636 CGCGGACAGCCATTGAATCTCCCTGGTATGTCGGATGCCCTGCGACACTACTGCT 1695  
Qy 360 SerLeuThrLeuValAsnAspLeuAlaIleThrAlaProAsnGlyThrLysThr 379  
Db 1696 TCTGTAAACCTCGTCAATGATTGGATTTGGTCAATCAGACCAACCAAGCAATAT 1755  
Qy 380 ValGlyAsnAspPheThrAlaProTyraAspAsnAsnTrpAspGlyArgAsnValGlu 399  
Db 1756 GTCGGGAATGACTTCTCAGACCAATTTGCAATATCTGGATGGCCGCAATACCTAGAA 1815  
Qy 400 AsnValPheIleAsnAlaProGlnSerGlyThrThrValGluValGlnAlaThrAsn 419  
Db 1816 AATGTATTATTAAATTCGCCCAAGTGGACATATACCATTTGAGGTGCAAGCATATAAT 1875  
Qy 420 ValProValSerProGlnThrPheSerLeuAlaIleValHis 433  
Db 1876 GTCCCGGTGGACCAAAACTTCTCGTTGGCAATTTGTAAC 1917

## RESULT 5

US-08-894-818B-2  
; Sequence 2, Application US/08894818B  
; Patent No. 6261822  
; GENERAL INFORMATION:  
; APPLICANT: TAKAKURA, Hikaru  
; APPLICANT: MORISHITA, Mio  
; APPLICANT: YAMAMOTO, Katsuhiko  
; APPLICANT: MITTA, Masanori  
; APPLICANT: ASADA, Kiyozo  
; APPLICANT: TSUNASAWA, Susumu  
; APPLICANT: KATO, Ikunoshin  
; TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENES  
; NUMBER OF SEQUENCES: 42  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Browdy and Neimark

STREET: 419 Seventh Street N.W., Ste. 300  
CITY: Washington  
STATE: D.C.  
COUNTRY: United States of America  
ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/894,818B  
FILING DATE: 20-MAY-1998  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/JP96/03253  
FILING DATE: 07-NOV-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 323285/1995  
FILING DATE: 12-DEC-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Browdy, Roger L.  
REGISTRATION NUMBER: 25,618  
REFERENCE/DOCKET NUMBER: TAKAKURA=1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 628-5197  
TELEFAX: (202) 737-3528  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1977 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: genomic DNA  
US-08-894-818B-2

## Alignment Scores:

Pred. No.:	8,91e-35	Length:	1977
Score:	432.50	Matches:	134
Percent Similarity:	44.85%	Conservative:	62
Best Local Similarity:	30.66%	Mismatches:	156
Query Match:	19.21%	Indels:	85
DB:	3	Gaps:	16

US-09-985-689A-7 (1-433) X US-08-894-818B-2 (1-1977)

Qy	8	ValLysAlaAspValAlaGlnAsnAsnPheGlyLeuTy-GlyGlnGlyGlnIleValAla	27
Db	433	ATAGGGCGGATACCGCTCGTGAACCTCCCTCGGTACAGCGAAGCGGTGTGGTGTGCC	492
Qy	28	ValAlaAspThrGlyLeuAspThrGlyArgAspAspSerSerMetHisGluAlaPheArg	47
Db	493	ATCGTCGATACGGGTATAGACGGGAAC-----CACCCGATCTGAAG	534
Qy	48	GlyLysIleThrAlaLeuTyR---AlaLeuGlyArgThrAsnAsnAlaAsnAspProAsn	66
Db	535	GGCAAGTTCATAGCTGCTGACGCGCTCAACGGCAGGTGACCCCTACGATGACCG	594
Qy	67	GlyHisGlyThrHisValAlaGlySerValLeuGly-----AsnAlaThrAsn	82
Db	595	GGACACGGAAACCCACGTTGGGGTATCGTTCCCGAACCGGCGCGCTTAACCTCCAGTAC	654
Qy	83	LysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMet-----AspSerGly	100
Db	655	ATAGCGTCCGCCCGCGGGAAGCTCGTCGGGTCAAGGTCTCGGTCCGACCGGTTCG	714
Qy	101	GlyGlyLeuGlyGlyLeuProAlaAsnLeuGlnThrLeuPheSerGlnAlaTySerAla	120
Db	715	GGAAAGCGTCTCCACCATCATCGGGGTGTGACTGGGTCTGTCAGAACAGGACAGTAC	774
Qy	121	GlyAlaArgIleHisThrAsnSerTyrGlyAlaProValAsnGlyAlaTyThrAsp	140
Db	775	GGGATAGGCTCATCAACCTCTCCCTCGGCTCTCCCGAGAGCTCCGACGGAACCGACTCC	834





1093 ACCCGATAAAGCACTACTACACCAAGCCTCTCGAACCAGCATGGCCACCCGACGTT 1152  
QY  
Db  
261 AlaGly---AsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyValThrPro 279  
Db  
1153 TCGGGCGTGGCGGCATCTCTCCAGGCCAC-----CCG 1188  
QY  
Db  
280 Lys-----ProSerLeuLeuLysAlaAlaLeuLleAlaGlyAla----- 292  
Db  
1189 AGCTGGACCCCGGCAAGAGTGAAGCCGCCCTCTATCGAGACCGCCGACATAGTCGCCGCC 1248  
QY  
Db  
293 -----AlaAspValGlyLeuGlyPheProAsnGlyAsnGlnGlyTyrGlyValGlyVal 309  
Db  
1249 AAGGAGATAGCGGACATCGCTACGTTGCG-----GATAGGGTG 1287  
QY  
Db  
310 ThrLeuAspLysSerLeu-----AsnValAlaPheValAsnGluThrSerProLeu 326  
Db  
1288 AACGTCTCAAGAGCCATCAAGTACGACGACTACGCCAAGCTCACCTTACCGGCTCCGTC 1347  
QY  
Db  
327 SerThrSerGlnLysAlaThrTyrSerPheThrAlaGlnAlaGlyLysProLeuLysIle 346  
Db  
1348 GCGGACAAAGGAAGCGCCACCCACCTTC----- 1377  
QY  
Db  
347 SerLeuValTyrSerAspAlaProGlySerThrThrAlaSerLeuThrLeuVal----- 364  
Db  
1378 -----GAGCTCAGCGCGCCACCTTCGTGACCCGCCACCTCTACTCGGAC 1422  
QY  
Db  
365 -----AsnAspLeuAspLeuValIleThrAlaProAsnGlyThrLysTyrValGly 381  
Db  
1423 ACGGGCTCGAGGACATCGACCTCTACCTCTACGACCCCAACGGGAACGAG---GTTGAC 1479  
QY  
Db  
382 AsnAspPheThrAlaProTyrAspAsnAsnTyrAspGlyArgAsnAsnValGluAsnVal 401  
Db  
1480 TACTCTACACCGCTACTAC-----GGCTTCGAGAAGGTC 1515  
QY  
Db  
402 PheIleAsnAlaProGlnSerGlyThrTyrThrValGluValGlnAlaTyr 418  
Db  
1516 GGCTACTACACCCGACCGCGGAACCTGACGCGTCAAGTCTGTCAGCTAC 1566

## RESULT 7

US-09-445-472-2  
; Sequence 2, Application US/09445472  
; Patent No. 6358726  
; GENERAL INFORMATION:  
; APPLICANT: TAKAKURA, Hikaru  
; APPLICANT: MORISHITA, Mio  
; APPLICANT: SHIMONO, Tomoko  
; APPLICANT: ASADA, Kiyozo  
; APPLICANT: KATO, Ikunoshin  
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE  
; FILE REFERENCE: TAKAKURA=6  
; CURRENT APPLICATION NUMBER: 1999-12-06  
; PRIOR FILING DATE: 1999-12-06  
; PRIOR APPLICATION NUMBER: 151969/1997  
; NUMBER OF SEQ ID NOS: 33  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 2  
; LENGTH: 1236  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Synthetic  
US-09-445-472-2

## Alignment Scores:

Pred. No.:	2,56-31	Length:	1236
Score:	396.00	Matches:	127
Percent Similarity:	42.56%	Conservative:	56
Best Local Similarity:	29.53%	Mismatches:	163
Query Match:	17.58%	Indels:	84
DB:	4	Gaps:	15

US-09-985-689A-7 (1-433) x US-09-445-472-2 (1-1236)

QY 16 AsnPheGlyLeuTyrGlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThr 35  
Db  
58 AACTTGGATATGATGTTCTGGAATCAACATAGGAATATTGACACTGGATTCAC--- 114  
QY 36 GlyArgAsnAspSerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAla 55  
Db 115 ---GCTTCTCATCCAGATCTCCAGGAAAGTA----- 144  
QY 56 LeuGlyArgThrAsnAsnAlaAsn-----AspProAsnGlyHisGly 69  
Db 145 ATTGGTGGGTAGATTTTTCATGTTAGGAGTTATCCATCATGATGACCATGGACATGA 204  
QY 70 ThrHisValAlaGlySerValLeuGly-----AsnAlaThrAsnLysGly 84  
Db 205 ACTCATCTAGCTTCAATAGCAGCTGGTGTCTGGAGCAAGTAATGGCAAGTACAGGGA 264  
QY 85 MetAlaProGlnAlaAsnLeuValPheGlnSerIleMet-----AspSerGlyGlyGly 102  
Db 265 ATGGCTCAGGAGTAAAGTGGCGGGAATTAAGTTCTAGTGCCTGCTGCTGGAAGC 324  
QY 103 LeuGlyGlyLeuProAlaAsnLeuGlnThrLeuPheSerGlnAlaTyrSerAlaGlyAla 122  
Db 325 ATATCTACTATAATTAAAGGAGTTGAGTGGCGGCTTGATAACAAAGATAAGTACGGAAT 384  
QY 123 ArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyValaTyrThrThrAspSerArg 142  
Db 385 AAGTCTAATCTTCTCTCTGTTTCAAGCCAGAGCTCAGATGGTACTGACGCTCTAAGT 444  
QY 143 AsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAlaGlyAsnGlu 162  
Db 445 CAGGCTCTTAATCAGCGTGGGATGCTGGATTAGTTGTTGTTGCTGCCCTGGAAACGT 504  
QY 163 GlyProGlySerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrValGly 182  
Db 505 GGACCTAACAGTATCAATCGGTTCTCCAGCAGCTGCAAGCAAGATTATTACAGTTGA 564  
QY 183 AlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsnHisValAla 202  
Db 565 GCC-----GTTGACAAGTATGATGTTATAACA 591  
QY 203 GlnPheSerSerArgGlyProThrArgAspGlyValIleLysProAspValMetAlaPro 222  
Db 592 AGCTTCTCAAGCAGAGGCGCAACTCGACAGCGGAGCTTAAGCTGAGTGTCTGCTCA 651  
QY 223 GlyThrTyrIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrpAlaAsn 242  
Db 652 GGAACCTGGATAATTCTGCCAGAGCAAGT-----GGAAGTACGATGGTCAACCA 702  
QY 243 HisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIleValAlaGly 262  
Db 703 ATTAATGACTATTACACAGCAGCTCTCGGACATCAATGGCACTCTCAGCTAGCTGT 762  
QY 263 AsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyValThrProLys----- 280  
Db 763 ATTGCAGCCCTCTTGTCTCCAA-----GCACACCCGAGCTGGACT 801  
QY 281 ProSerLeuLeuLysAlaAlaLeuLleAlaGlyAla----- 292  
Db 802 CCAGACAAAGTAAACAGCCCTCATAGAACTGCTGATATCGTAAAGCCAGATGAATA 861  
QY 293 AlaAspValGlyLeuGlyPheProAsnGlyAsnGlnGlyTyrGlyArgValThrLeuAsp 312  
Db 862 GCGGATATAGCTACGTTGCA-----GGTAGGGTTTAAATGCATAC 900  
QY 313 LysSerLeu-----AsnValAlaPheValAsnGluThrSerProLeuSerThrSer 329  
Db 901 AAGCTATAACTACGATAAATATGCAAGCTAGTGTCTCAGTGGATATGTTGCCACAAA 960  
QY 330 GlnLysAlaThrTyrSerPheThrAlaGlnAlaGlyLysProLeuLysIleSerLeuVal 349  
Db 961 GGCAGCCAAACTCACCAGTTCTGTTATTAGCGGAGCTTCGTTAAGTAACTGCCACATTATAC 1020

Qy 350 TipSexAspAlaProGlySerThrThrAlaSerLeuThrLeuValAsnAspLeuAspLeu 369  
Db 1021 TGGACAATGCCAAT-----AGGACCTTGACTT 1050  
Qy 370 ValIleThrAlaProGlySerThrThrAlaSerLeuValAsnAspLeuAspLeu 389  
Db 1051 TACCTCTAGATCCCAATGGAACCAAG---GTGACTACTCTTACACCGCTACTAT--- 1104  
Qy 390 AsnAsnTrpAspGlyArgAsnValGluAsnValPheIleAsnAlaProGlnSerGly 409  
Db 1105 -----GGATTCGAAAGAGTTGGTTATTACAAACCACTGATGGA 1143  
Qy 410 ThrThrValGluValGlnAlaValAsn 419  
Db 1144 ACATGGACAATTAAAGTTGTGAAGCTACAGC 1173

## RESULT 8

US-08-894-818B-4  
Sequence 4, Application US/08894818B  
Patent No. 6261822  
GENERAL INFORMATION:  
APPLICANT: TAKAKURA, Hikaru  
APPLICANT: MORISHITA, Mio  
APPLICANT: YAMAMOTO, Katsuhiko  
APPLICANT: MITTA, Masanori  
APPLICANT: ASADA, Kiyozo  
APPLICANT: TSUNASAWA, Susumu  
APPLICANT: KATO, Ikunoshin  
TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENES  
NUMBER OF SEQUENCES: 42  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Broadway and Neimark  
STREET: 419 Seventh Street N.W., Ste. 300  
CITY: Washington  
STATE: D.C.  
COUNTRY: United States of America  
ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/894,818B  
FILING DATE: 20-MAY-1998  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/JP96/03253  
FILING DATE: 07-NOV-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 323285/1995  
FILING DATE: 12-DEC-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Browdy, Roger L.  
REGISTRATION NUMBER: 25,618  
REFERENCE/DOCKET NUMBER: TAKAKURA=1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 628-5197  
TELEFAX: (202) 737-3528  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1566 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: genomic DNA  
FEATURE:  
OTHER INFORMATION: /note= N at position 1283 is G or T.  
US-08-894-818B-4

Alignment Scores:

Pred. No.: 3,63e-31 Length: 1566

Score: 396.00 Matches: 127  
Percent Similarity: 42.56% Conservative: 56  
Best Local Similarity: 29.53% Mismatches: 163  
Query Match: 17.58% Indels: 84  
DB: 3 Gaps: 15

US-09-985-689A-7 (1-433) x US-08-894-818B-4 (1-1566)

Qy 16 AsnPhgGlyLeuTyrGlyGlnGlyValAlaValAlaAspThrGlyLeuAspThr 35  
Db 58 AACTTGGGATATGATGTTCTGGAATCAATAGCAATAATTGACACTGCAATTGAC--- 114  
Qy 36 GlyArgAsnAspSerSerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAla 55  
Db 115 -----GCTTCTCATCCAGATCTCCAGGAAAGTA----- 144  
Qy 56 LeuGlyArgThrAsnAsnAlaAsn-----AspProAsnGlyHisGly 69  
Db 145 ATTGGGTGGTAGATTTTGTCAATGGTAGGAGTTATCCATACGATGACCATGACATGGA 204  
Qy 70 ThrHisValAlaGlySerValLeuGly-----AsnAlaThrAsnLysGly 84  
Db 205 ACTCATGTAGCTTCAATAGACAGCTGGTACTGGACAGCAAGTAATGGCAAGTACAGGGA 264  
Qy 85 MetAlaProGlnAlaAsnLeuValPheGlnSerIleMet-----AspSerGlyGlyGly 102  
Db 265 ATGGCTCCAGGAGCTTAAGCTGGCGGAATTAAGGTTCTAGGTGCGGATGTTCTTGAAGC 324  
Qy 103 LeuGlyGlyLeuProAlaAsnLeuGlnThrLeuPheSerGlnAlaTyrSerAlaGlyAla 122  
Db 325 ATATCTACTATAATTAAGGGAGTTGAGTGGCGCTTGTATACAAAGATAAGTAGCGAATT 384  
Qy 123 ArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyrThrAspSerArg 142  
Db 385 AAGGTCAATTAATCTTCTCTGGTTCAAGCCAGAGCTCAGATGCTACTGACGCTCTAAGT 444  
Qy 143 AsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAlaGlyAsnGlu 162  
Db 445 CAGGCTGTATTCAGCGTGGATGCTGGATTAGTTGTTGCTGCGCTGGAACAGT 504  
Qy 163 GlyProGlySerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrValGly 182  
Db 505 GGACCTAAACAAGTATACAAATCGTTCTCCAGCAGCTGCAAGCAAAAGTTATTACAGTTGA 564  
Qy 183 AlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsnHisValAla 202  
Db 565 GCC-----GTTGACAGTATGATGTATAACA 591  
Qy 203 GlnPheSerSerArgGlyProThrArgAspGlyArgIleLysProAspValMetAlaPro 222  
Db 592 AGCTTCTCAAGCAGAGGCGCAACTGCAGCGCAGGCTTAAGCCTGAGGTTGTTGCTCCA 651  
Qy 223 GlyThrTyrIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrpAlaAsn 242  
Db 652 GGAAACTGGATAATTGCTGCCAGGCAAGT-----GGAACTAGCATGGGTCAACCA 702  
Qy 243 HisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIleValAlaGly 262  
Db 703 ATTAATGACTATTACACAGCAGCTCCTGGGACATCAATGGCACTCTCTCAGTAGCTGCT 762  
Qy 263 AsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyValThrProLys----- 280  
Db 763 ATTCAGGCGCTCTTGTCTCAA-----GCACACCGAGCTGGACT 801  
Qy 281 ProSerLeuLeuLysAlaAlaLeuIleAlaGlyAla----- 292  
Db 802 CCAGACAAAGTAAACACAGCCCTCATAGAAACTCTGATATCGTTAAAGCCAGATGAATA 861  
Qy 293 AlaAspValGlyLeuGlyPheProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAsp 312  
Db 862 GCCGATATAGCTACGGTGCA-----GGTAGGTTATGCTATC 900  
Qy 313 LysSerLeu-----AsnValAlaPheValAsnGluThrSerProLeuSerThrSer 329

Db 901 AAGGCTATAAAGTACGTAAGTATCAAGAGTGGTTTCAGTGGATGTTGGCAACAAA 960  
QY 330 GlnLysAlaThrTySerPheThrAlaGlnAlaGlyLysProLeuLysIleSerLeuVal 349  
Db 961 GCAGCGCAACATCACAGTTCGTTATTAGCGAGCTTCGTTGTAAGTGCACATATATAC 1020  
QY 350 TrpSerAspAlaProGlySerThrThrAlaSerLeuThrLeuValAsnAspLeuAspLeu 369  
Db 1021 TGGGCAATGCCAAT-----AGCGACCTTGATCTT 1050  
QY 370 ValIleThrAlaProAsnGlyThrLysTyrValGlyAsnAspPheThrAlaProTyrAsp 389  
Db 1051 TACCTCTACGATCCCAATGAACAG---GTTCAGTACTCTTACACCGCTACTAT--- 1104  
QY 390 AsnAsnTrpAspGlyArgAsnAsnValGluAsnValPheIleAsnAlaProGlnSerGly 409  
Db 1105 -----GGATTCGAAAGGTTGGTTATTACAAACCACTGATGGA 1143  
QY 410 ThrTyrThrValGluValGlnAlaTyrAsn 419  
Db 1144 ACATGGACAATTAAGTTGTAAGCTACAGC 1173

## RESULT 9

US-08-894-818B-34  
Sequence 34, Application US/08894818B  
Patent No. 6261822  
GENERAL INFORMATION:  
APPLICANT: TAKAKURA, Hikaru  
APPLICANT: MORISHITA, Mio  
APPLICANT: YAMAMOTO, Katsuhiko  
APPLICANT: MITTA, Masanori  
APPLICANT: ASADA, Kiyozo  
APPLICANT: TSUNASAWA, Susumu  
APPLICANT: KATO, Ikunoshin  
TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENES  
NUMBER OF SEQUENCES: 42  
CORRESPONDENCE ADDRESS:  
ADDRESSER: Browdy and Neimark  
STREET: 419 Seventh Street N.W., Ste. 300  
CITY: Washington  
STATE: D.C.  
COUNTRY: United States of America  
ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: IBM PC compatible  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
FILING DATE: 20-MAY-1998  
PRIORITY DATE: 07-NOV-1996  
CLASSIFICATION: 435  
APPLICATION NUMBER: PCT/JP96/03253  
FILING DATE: 07-NOV-1996  
PRIORITY DATE: 07-NOV-1996  
APPLICATION NUMBER: JP 323285/1995  
FILING DATE: 12-DEC-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Browdy, Roger L.  
REGISTRATION NUMBER: 25,618  
REFERENCE/DOCKET NUMBER: TAKAKURA=1  
TELEPHONE: (202) 528-5197  
TELEFAX: (202) 737-3528  
INFORMATION FOR SEQ ID NO: 34:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1962 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: genomic DNA

US-08-894-818B-34  
Alignment Scores:  
Pred. No.: 5,19e-31 Length: 1962  
Score: 396.00 Matches: 127  
Percent Similarity: 42.56% Conservative: 56  
Best Local Similarity: 29.53% Mismatches: 163  
Query Match: 17.58% Indels: 84  
DB: 3 Gaps: 15  
US-09-985-689A-7 (1-433) x US-08-894-818B-34 (1-1962)  
QY 16 AsnPheGlyLeuTyrGlyGlnIleValAlaValAlaAspThrGlyLeuAspThr 35  
Db 454 AACITGGGATATGATGGTTCTGGATCACAATAGGATAATTCACACTGGAATTGAC--- 510  
QY 36 GlyArgAsnAspSerSerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAla 55  
Db 511 -----GCTTCTCATCCAGATCTCCAAGGAAAAGTA----- 540  
QY 56 LeuGlyArgThrAsnAsnAlaAsn-----AspProAsnGlyHisGly 69  
Db 541 ATTGGGTGGTAGATTTTGCAATGGTAGGATTCATCATGATGACCATGACATGACATGGA 600  
QY 70 ThrHisValAlaGlySerValLeuGly-----AsnAlaThrAsnLysGly 84  
Db 601 ACTCATGTAGTCTCAATAGCAGCTGGTACTGGAGCAGCAAGTAATGGCAAGTAAAGGGA 660  
QY 85 MetAlaProGlnAlaAsnLeuValPheGlnSerIleMet-----AspSerGlyGlyGly 102  
Db 651 ATGGCTCCAGGAGCTAAGCTGGCGGAATTAAGGTTCTAGGTGCGATGTTCTTGGGAAGC 720  
QY 103 LeuGlyGlyLeuProAlaAsnLeuGlnThrLeuPheSerGlnAlaTyrSerAlaGlyAla 122  
Db 721 ATATCTACTATAATTAAGGGAGTTGAGTGGGCGCTTGCATTAACAAAGATAAGTACGGAAT 780  
QY 123 ArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyrThrThrAspSerArg 142  
Db 781 AAGGTCAATTAATCTTCTCTTGGTTCAAGCCAGAGCTCAGATGCTACTGACGCTTAAGT 840  
QY 143 AsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaGlyAsnGlu 162  
Db 841 CAGGCTGTAAATGAGCGGTGGGATGCTGGATTAAGTCTGTTGCTGCTGCGCTGGAACAGT 900  
QY 163 GlyProGlySerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrValGly 182  
Db 901 GGACCTAACAAAGTATACAACTCGGTTCTCCAGCAGCTGCAAGCAAGATTAATACAGTTGA 960  
QY 183 AlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsnHisValAla 202  
Db 961 GCC-----GTTGACAAGTATGATGTTATAACA 987  
QY 203 GlnPheSerSerArgGlyProThrArgAspGlyArgIleLysProAspValMetAlaPro 222  
Db 988 AGCTTCTCAAGCAGAGCGGCCCAACTGCAGACGCGAGGCTTAAGCTGAGGTTGTTGCTCCA 1047  
QY 223 GlyThrTyrIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrpAlaAsn 242  
Db 1048 GGAAACTGGATAATGTCTGCCAGAGCAAGT-----GGAACCTAGCTAGGTTGCTGCTCA 1098  
QY 243 HisAspSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProIleValAlaGly 262  
Db 1099 ATTAATGACTATTACACAGCAGCTCTCTGGGACATCAATGCAACTCTCTCAGTAGCTGGT 1158  
QY 263 AsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyValThrProLys----- 280  
Db 1159 ATTGAGCGCTCTTGTCTCAA-----GCACACCCGAGCTGGAAT 1197  
QY 281 ProSerLeuLeuLysAlaAlaLeuIleAlaGlyAla----- 292  
Db 1198 CCAGACAAGTAATAAACAGCCCTCATAGAAACTGCTGATATCGTAAAGCCAGATGAAATA 1257  
QY 293 AlaAspValGlyLeuGlyPheProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAsp 312

Db 1258 GCCGATATAGCTACGGTGC-----GTTAGGGTTAATGCATAC 1296  
Qy 313 LysSerLeu-----AsnValAlaPheValAsnGluThrSerProLeuSerThrSer 329  
Db 1297 AAGGCTATAAACTACGATAACTATGCAAGAGTAGTGTCTCACTGGATATGTTGCCAACAAA 1356  
Qy 330 GlnLysAlaThrTySerPheThrAlaGlnAlaGlyLysProLeuLysLysSerLeuVal 349  
Db 1357 GGCAGCAAACTCACCAAGTTCGTTATTAGCGGAGCTTCGTTCACTGCCCATATATAC 1416  
Qy 350 TrpSerAspAlaProGlySerThrThrAlaSerLeuThrLeuValAsnAspLeuAspLeu 369  
Db 1417 TGGGACAAATGCCAAT-----AGCCACCTTGATCTT 1446  
Qy 370 ValIleThrAlaProAsnGlyThrLysTy-ValGlyAsnAspPheThrAlaProTyAsp 389  
Db 1447 TACCTCTACGATCCCAATGGAACACAG---GTTGACTACTCTTACACCGCTACTAT--- 1500  
Qy 390 AsnAsnTrpAspGlyA:GAsnAsnValGluAsnValPheIleAsnAlaProGlnSerGly 409  
Db 1501 -----GGATTGAAAAGGTTGGTTATTACACCACTGATGGA 1539  
Qy 410 ThrTyThrValGluValGlnAlaTy:Asn 419  
Db 1540 ACATGGACAATTAAGTTGTAAGCTACAGC 1569

## RESULT 10

US-09-445-472-15  
; Sequence 15, Application US/09445472  
; Patent No. 6358726  
; GENERAL INFORMATION:  
; APPLICANT: TAKAKURA, Hikaru  
; APPLICANT: MORISHITA, Mio  
; APPLICANT: SHIMOJO, Tomoko  
; APPLICANT: ASADA, Kiyoza  
; APPLICANT: KATO, Ikunoshi  
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE  
; FILE REFERENCE: TAKAKURA-6  
; CURRENT APPLICATION NUMBER: US/09/445,472  
; CURRENT FILING DATE: 1999-12-06  
; PRIOR APPLICATION NUMBER: 151969/1997  
; PRIOR FILING DATE: 1997-06-10  
; NUMBER OF SEQ ID NOS: 33  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 15  
; LENGTH: 1962  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Synthetic  
US-09-445-472-15

Alignment Scores:  
Pred. No.: 5,19e-31 Length: 1962  
Score: 396.00 Matches: 127  
Percent Similarity: 42.56% Conservative: 56  
Best Local Similarity: 29.53% Mismatches: 163  
Query Match: 17.58% Indels: 84  
DB: 4 Gaps: 15

US-09-985-689A-7 (1-433) x US-09-445-472-15 (1-1962)

Qy 16 AsnPheGlyLeuTy:GlyGlnGlyGlnIleValAlaAlaAspThrGlyLeuAspThr 35  
Db 454 AACTGGGATATGATGGTTCTCGAATCAATAGCAATAGCAATAGCAATGATGAC--- 510  
Qy 36 GlyArgAsnAspSerSerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyAla 55  
Db 511 -----GCTTCTCATCCAGATCTCCCAAGGAAAGTA----- 540  
Qy 56 LeuGlyArgThrAsnAsnAlaAsn-----AspProAsnGlyHisGly 69  
Db 511 -----GGATTGAAAAGGTTGGTTATTACACCACTGATGGA 1539

Db 541 ATTGGTGGGTAGATTTTGTCAATGGTAGGAGTTATCCATACGATGACCATGGACATGGA 600  
Qy 70 ThrHisValAlaGlySerValLeuGly-----AsnAlaThrAsnLysGly 84  
Db 601 ACTCATGTAGTCTCAATAGCAGCTGGTACTGGAGCAGCAAGTAATGCGAAGTACAGGGA 660  
Qy 85 MetAlaProGlnAlaAsnLeuValPheGlnSerIleMet-----AspSerGlyGlyGly 102  
Db 661 ATGGCTCCAGGAGCTAAGCTGGCGGAATTAAAGTTCTTAGGTGCGGATGTTCTTGGGAGC 720  
Qy 103 LeuGlyGlyLeuProAlaAsnLeuGlnThrLeuPheSerGlnAlaTySerAlaGlyAla 122  
Db 721 ATATCTACTATATAATTAAGGAGTTGAGTGGGCGGCTTGTATACAAAGATAAGTACGGAAT 780  
Qy 123 ArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyThrThrAspSerArg 142  
Db 781 AAGGTCAATTAATCTTCTCTTGGTTCAAGCCAGAGCTCAGATGGTACTGACGCTCTAAGT 840  
Qy 143 AsnValAspAspTyValArgLysAsnAspMetThrIleLeuPheAlaAlaGlyAsnGlu 162  
Db 841 CAGGCTGTTAATGCGAGCGTGGGATGCTGCAATAGTTGTTGGTTGCGCTGGGAAACAGT 900  
Qy 163 GlyProGlySerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrValGly 182  
Db 901 GGACCTACACAGTATACATCGTTCTCCAGCAGCTGCAGCAAGATGTTATTACAGTTGGA 960  
Qy 183 AlaThrGluAsnLeuArgProSerPheGlySerTyAlaAspAsnIleAsnHisValAla 202  
Db 961 GCC-----GTTGCAAGTATGATGTTATAACA 987  
Qy 203 GlnPheSerArgGlyProThrArgAspGlyArgIleLysProAspValMetAlaPro 222  
Db 988 AGCTTCTCAAGCAGAGGCGCAACTGCACAGCGAGCTTAGCTGAGTGTGTTGCTCCA 1047  
Qy 223 GlyThrTyIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrpAlaAsn 242  
Db 1048 GGAACCTGGATAATGCTGCCAGACAAAGT-----GGAACCTAGCATGGGTCAACCA 1098  
Qy 243 HisAspSerLysTy:AlaTyMetGlyGlyThrSerMetAlaThrProIleValAlaGly 262  
Db 1099 ATTAATGACTATTACACAGCAGCTCTCGGACATCAATGGCACTTCCTCAGTACTGGT 1158  
Qy 263 AsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyValThrProLys----- 280  
Db 1159 ATTGCAGCCCTCTTGCTCCAA-----GCACACCCGAGCTGGACT 1197  
Qy 281 ProSerLeuLysAlaAlaLeuIleAlaGlyAla----- 292  
Db 1198 CCAGACAAAGTAAACAGCCCTCATAGAACTGCTGATATCGTAAAGCCAGATGAATA 1257  
Qy 293 AlaAspValGlyLeuGlyPheProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAsp 312  
Db 1258 GCCGATATAGCTACGGTCCA-----GGTAGGGTTAATGCATAC 1296  
Qy 313 LysSerLeu-----AsnValAlaPheValAsnGluThrSerProLeuSerThrSer 329  
Db 1297 AAGGCTATAAACTACGATAACTATGCAAGAGTAGTGTTCACCTGGATATGTTGCCAACAAA 1356  
Qy 330 GlnLysAlaThrTySerPheThrAlaGlnAlaGlyLysProLeuLysLysSerLeuVal 349  
Db 1357 GGCAGCAAACTCACCAAGTTCGTTATTAGCGGAGCTTCGTTCACTGCCCATATATAC 1416  
Qy 350 TrpSerAspAlaProGlySerThrThrAlaSerLeuThrLeuValAsnAspLeuAspLeu 369  
Db 1417 TGGGACAAATGCCAAT-----AGCCACCTTGATCTT 1446  
Qy 370 ValIleThrAlaProAsnGlyThrLysTyValGlyAsnAspPheThrAlaProTyAsp 389  
Db 1447 TACCTCTACGATCCCAATGGAACACAG---GTTGACTACTCTTACACCGCTACTAT--- 1500  
Qy 390 AsnAsnTrpAspGlyArgAsnAsnValGluAsnValPheIleAsnAlaProGlnSerGly 409  
Db 1501 -----GGATTGAAAAGGTTGGTTATTACACCACTGATGGA 1539



QY 384 PheThrAlaProTyrAspAsnAsnTyrPaspGlyArgAsnAsnValGluAsnValPheIle 403  
Db 1486 TACACCGCTACTAC-----GGTTCCAGAGAGTCCGCTAC 1521  
QY 404 AsnAlaProGlnSerGlyThrTyrThrValGluValGlnAlaTyr 418  
Db 1522 TACAACCCGACCCCGGAACCTGACGGTCAAGGTCGTCTAGCTAC 1566

RESULT 12  
US-09-000-016-3  
; Sequence 3, Application US/09000016  
; Patent No. 6143541  
; GENERAL INFORMATION:  
; APPLICANT: Akira ARISAWA et al.  
; TITLE OF INVENTION: GENE ENCODING A PROTEIN HAVING ASYMMETRIC  
; TITLE OF INVENTION: HYDROLASE ACTIVITY FOR 4-SUBSTITUTED 1,4-DIHYDROPYRIDINE DERIV  
; TITLE OF INVENTION: ITS EXPRESSION PRODUCT  
; NUMBER OF SEQUENCES: 7  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.  
; STREET: 2033 K Street, N.W., #800  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: U.S.A.  
; ZIP: 20006  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: MS-DOS  
; SOFTWARE: Wordperfect 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/000,016  
; FILING DATE: January 30, 1998  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Warren M. Cheek, Jr.  
; REGISTRATION NUMBER: 33,367  
; REFERENCE/DOCKET NUMBER:  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-721-8200  
; TELEFAX: 202-721-8250  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2539 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: genomic DNA  
; ORIGINAL SOURCE:  
; ORGANISM: Streptomyces viridosporus  
; STRAIN: A-914  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 338...2539  
; IDENTIFICATION METHOD: E  
US-09-000-016-3

Alignment Scores:  
Pred. No.: 2,84e-24 Length: 2539  
Score: 332.50 Matches: 137  
Percent Similarity: 41.05% Conservative: 51  
Best Local Similarity: 29.91% Mismatches: 159  
Query Match: 14.76% Indels: 111  
DB: 3 Gaps: 23

US-09-985-689A-7 (1-433) x US-09-000-016-3 (1-2539)

QY 18 GlyLeuTyrGlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArg 37

Db 998 GGCTACACGGCAGGCGCTGAAGATCCGCTCTGGACACCGGTCTCGACACGAGC--- 1054  
QY 38 AsnAspSerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGly 57  
Db 1055 -----CATCCGGACCTGAGGGCGGGTGACCGCGTCCAAAGAACTTCACC 1099  
QY 58 ArgThrAsnAsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeu 77  
Db 1100 GCCCGCGCGCGCGCGGCAAGGTGGGTCACGGCACCCACCTCGCTCGATCGCGGCG 1159  
QY 78 GlyAsnAlaThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuVal 92  
Db 1160 GGCAACGGCGCGCGCGCGCAAGGTGCAAGGGCGCTCGCACCGCGCGCGGATCCTC 1219  
QY 93 PheGlnSerIleMet---AspSerGly---GlyGlyLeuGlyGlyLeuProAlaAsnLeu 110  
Db 1220 AACGGCAAGGTCCTCGACGACTCCGGTTTCGGGACACCTCCGGCATCTCCGCGCATG 1279  
QY 111 GlnThrLeuPheSerGlnAlaTyrSerAlaGlyAlaArgIleHisThrAsnSerTyrGly 130  
Db 1280 CAGTGGCGCGCGCGCGAG-----GGCGCGACGCTCGTCACCATGAGCTGGGC 1327  
QY 131 Ala-----ProValAsnGlyAlaTyrThrThrAspSerArgAsn 143  
Db 1328 GGCATGGACACACCGGAGACCGACCCGCTGGAGCGGCG----- 1366  
QY 144 ValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAlaGlyAsnGly 163  
Db 1367 CTCGACAAGCTGTCGCGGAGAGGGGCTCTGTCGTCATCGCGCGCGCAACGAGGCG 1426  
QY 164 ProGlySerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrValGlyAla 183  
Db 1427 CCGGAGTCCG-----ATCGGTTCCGCGGACGCGGACGCGCGCTCCACGTCGGCGCC 1480  
QY 184 ThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsnHisValAlaGln 203  
Db 1481 -----GTCGACGACAGGACAGCTCGCGAC 1507  
QY 204 PheSerSerArgGlyPro---ThrArgAspGlyArgIleLysProAspValMerAlaPro 222  
Db 1508 TTCTCTCTCACCGCGCGCTCGCGACGCGCCCATCAAGCGCGGAGCTCACCCTCC 1567  
QY 223 GlyThrTyrIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTyrAlaAsn 242  
Db 1568 GCGGTGGACATCACGCGCGCTCGCGGAGGCGGACGACATCGCGCAGGAGGTGGTGAG 1627  
QY 243 HisAspSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProIleValAlaGly 262  
Db 1628 GGACCGCGCGGCTACATGACCATCTCCGGCACGTCGATGGGACCCCGCACGTCGCGGCG 1687  
QY 263 AsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyValThrProLysProSer 282  
Db 1688 GCGCGCGCGCTCTCTGAAGCAGCAG-----CACCCCGAC 1720  
QY 283 LeuLeuLysAlaAlaLeu-----IleAlaGlyAlaAlaAspValGlyLeuGlyPhe 299  
Db 1721 TGGACCTCGCGCGCAACTGAAGGCGCGCTCACCGGCTCCACCAAGGCGGCAAGTACACC 1780  
QY 300 ProAsnGlyAsnGlnGlyTyrPheGlyArgValThrLeuAspLysSerLeuAsnValAlaPhe 319  
Db 1781 CCG---TTCGACGAGGGTTCCGGCGCGATCCAGGCGGCAAGGCGCTCCAGCAGACCGGTG 1837  
QY 320 ValAsnGluThrSerProLeuSerThrSer----- 329  
Db 1838 ATC-----GCCGACCGCGTCTCGGTGAGCTTCGGGCGTCCAGCAGTGGCGGCACACCGAC 1891  
QY 330 -----GlnLysAlaThrTyrSerPheThrAlaGlnAlaGlyLysProLeu 344  
Db 1892 GACGAGCGGTCACCAAGCAGCTGACCTACCGCAACCTCGGCACCCACGAGCATCGCTG 1951  
QY 345 LysIleSerLeuValTyrSerAsp-----AlaPro----- 354







```
QY 320 ValAsnGluThrSerProLeuSerThrSer----- 329
Db 1838 ATC-----GCCACCGGCTCTCGTGAGCTTCGGGCTCCAGAGTGGCGGCACACCGAC 1891
QY 330 -----GlnLysAlaThrTyrsrPheThrAlaGlnAlaGlyLysProLeu 344
Db 1892 GAGCAGCGGTCACCAAGCAGCTGACCTACCGCAAGCTCGGCACCCAGGACGTCAGCTG 1951
QY 345 LysIleSerLeuValTrpSerAsp-----AlaPro----- 354
Db 1952 AAGCTGACGTGACCGCCACCGACCCCAAGGCAAGGGCAAGCGGCGGCGGCTTCTTCAGC 2011
QY 355 ---GlySerThrThrAlaSerLeuThrLeuValAsnAspLeuValIleThrAla 373
Db 2012 CTGGCGCCACCCAGC-----GTGACCGTC 2035
QY 374 ProAsnGlyThrLysTyrsrValGlyAsnAspPheThrAlaProTyrsrAspAsnTrpAsp 393
Db 2036 CCGCGCGCGGCGAGC---GCCTCGTCGACATACCGCC-----GACACCGGCTCGGC 2086
QY 394 GlyArg---AsnAsnValGluAsnValPheIleAsnAlaProGlnSerGlyThrTyThr 412
Db 2087 GGCACGGTGAGCGCGGTACTCGGCTACGTGTGTCGC---ACGGCGCGCGGCGAGCG 2143
QY 413 Val-----GluValGlnAlaTyrsrAsnValProVal 422
Db 2144 GTCGCGACGCGCGCGCGGTGCAGCGGAGTGCAGTCGACGAGTGCACCGTC 2197

RESULT 14
US-09-000-016-1
; Sequence 1, Application US/09000016
; Patent No. 6143541
; GENERAL INFORMATION:
; APPLICANT: Akira ARISAWA et al.
; TITLE OF INVENTION: GENE ENCODING A PROTEIN HAVING ASYMMETRIC
; TITLE OF INVENTION: HYDROLASE ACTIVITY FOR 4-SUBSTITUTED 1,4-DIHYDROPYRIDINE DERIV
; TITLE OF INVENTION: ITS EXPRESSION PRODUCT
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.
; STREET: 2033 K Street, N.W., #800
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20006
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/000,016
; FILING DATE: January 30, 1998
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Warren M. Cheek, Jr.
; REGISTRATION NUMBER: 33,367
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-721-8200
; TELEFAX: 202-721-8250
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2809 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
; ORIGINAL SOURCE:
```

```
ORGANISM: Streptomyces viridosporus
STRAIN: A-914
ORIGINAL SOURCE:
ORGANISM: Streptomyces antibioticus
STRAIN:
FEATURE:
NAME/KEY: CDS
LOCATION: 338...2539
IDENTIFICATION METHOD: E
FEATURE:
NAME/KEY: CDS
LOCATION: 2540...2809
IDENTIFICATION METHOD: P
US-09-000-016-1
Alignment Scores:
Pred. No.: 1,09e-23 Length: 2809
Score: 327.50 Matches: 136
Percent Similarity: 40.83% Conservative: 51
Best Local Similarity: 29.69% Mismatches: 160
Query Match: 14.54% Indels: 111
DB: 3 Gaps: 23
US-09-985-689a-7 (1-433) x US-09-000-016-1 (1-2809)
QY 18 GlyLeuTyGlyGlnIleValAlaValAlaAsePThrGlyLeuAspThrGlyArg 37
Db 998 GGTACGACGCGGAGGGGTGAAGATCGCTCTGGACACCGGTGTGACACGAGC--- 1054
QY 38 AsnAspSerSerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyThrAlaLeuGly 57
Db 1055 -----CATCCGACCTGAAGGCGCGGTGACCGGTCCAGAACTTCACC 1099
QY 58 ArgThrAsnAsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeu 77
Db 1100 GCGCGCGCGCGCGCGGACAAAGTGGCGCACGACCGACCGCTCGCTCGATCGCGCG 1159
QY 78 GlyAsnAlaThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuVal 92
Db 1160 GGCACGGCGCGCCAGTCCCAAGGGCAAGTACAAGGGCGTCCGACCGCGCGCGATCTC 1219
QY 93 PheGlnSerIleMet---AspSerGly---GlyGlyLeuGlyGlyLeuProAlaAsnLeu 110
Db 1220 AACGGCAGGTCTCTGACGACTCCCGTTTCGGCGACGACTCGCGCATCTCTCGCGCATG 1279
QY 111 GlnThrLeuPheSerGlnAlaTyrsrAlaGlyAlaArgIleHisThrAsnSerTrpGly 130
Db 1280 GAGTGGCGCGCGCGCAG-----GGCGCGCGCGTCTCAACATGAGCTGGGCGC 1327
QY 131 Ala-----ProValAsnGlyAlaTyThrThrAspSerArgAsn 143
Db 1328 GGCATGGACACACCGGAGACCGCCGCTGGAGGCGCGG----- 1366
QY 144 ValAspAspTyThrValArgLysAsnAspMetThrIleLeuPheAlaAlaGlyAsnGly 163
Db 1367 GTCGACAAGCTGTCGCCGAGAGGGCGTCTTCGCTCGCATCGCGCGCGCAACGAGGCG 1426
QY 164 ProGlySerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrValGlyAla 183
Db 1427 CCGGAGTCG-----ATCGGTTTCGCCGCGCGCGCGCGCGCGCTCAACGTCGCGCGC 1480
QY 184 ThrGluAsnLeuArgProSerPheGlySerTyThrAlaAspAsnIleAsnHisValAlaGln 203
Db 1481 -----GTCGACGACAAAGGACAAAGTCCGCGAC 1507
QY 204 PheSerSerArgGlyPro---ThrArgAspGlyArgIleLysProAspValMetAlaPro 222
Db 1508 TTCTCTCCACCGCGCGCGCGCTCGCGCGACGCGCGCATCAAGCGCGGACGCTCCCGCTCC 1567
QY 223 GlyThrTyThrIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrpAlaAsn 242
Db 1568 GCGGTGGACATACGCGCGCGCTCGCGCGAGGGGCAACATCGCGCGAGGAGGTGCGTGAG 1627
```

QY 243 HisAspSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProIleValAlaGly 262  
Db 1628 GGACGGCGCGCTACATGACCATCTCCGGCAGCTGATGCGCAGCCCGCAGCTCGCGGGC 1687  
QY 263 AsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyValThrProLysProSer 282  
Db 1688 GCGGGCGCTCTCTGAAGCAGCAG-----CHCCCGCAG 1720  
QY 283 LeuLeuLysAlaAlaLeu-----IleAlaGlyAlaAlaAspValGlyLeuGlyPhe 299  
Db 1721 TGGACCTCCGGCGGACTGAAGCGCGGCTCACCAGCTCCACCAAGGCGCGCAATACACC 1780  
QY 300 ProAsnGlyAsnGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsnValAlaPhe 319  
Db 1781 CCG---TTCGAGCAGGTTCCGGCGGATCCAGGCCGCAAGGCGCTCCAGCAGACCGTG 1837  
QY 320 ValAsnGluThrSerProLeuSerThrSer----- 329  
Db 1838 ATC-----GCGACCCCGGTTCCGGTGAGCTTCGCGCTCCAGCAGTCCGCGCACCGCAG 1891  
QY 330 -----GlnLysAlaThrTyrSerPheThrAlaGlnAlaGlyLysProLeu 344  
Db 1892 GACGAGCGGTCACCAAGCAGCTGACCTACCGCAACCTCGGCACCCAGGACGTCAGCTG 1951  
QY 345 LysIleSerLeuValTrpSerAsp-----AlaPro----- 354  
Db 1952 AAGCTGAGCTGACCGCCACCGCAGCCAGCCAGGCGGCGCGCGGCTTCTTCACG 2011  
QY 355 ---GlySerThrThrAlaSerLeuThrLeuValAsnAspLeuAspLeuValIleThrAla 373  
Db 2012 CTGGCGGCCACCGC-----GTGACCGTC 2035  
QY 374 ProAsnGlyThrLysTyrValGlyAsnAspPheThrAlaProTyrAspAsnAsnTrpAsp 393  
Db 2036 CCGCGCGCGCGCAGC---GCTCCGCTGACATGACCGCC-----GACACCCGCGCTCGGC 2086  
QY 394 GlyArg---AsnAsnValGluAsnValPheIleAsnAlaProGlnSerGlyThrTyrThr 412  
Db 2087 GGCACGGTGGAGCGGCGTACTCGCGTACGTGTCGCC---ACGGCGCGCGCGCAGACG 2143  
QY 413 Val-----GluValGlnAlaTyrAsnValProVal 422  
Db 2144 GTCCGCGCGCGCGCGTGCAGCGCGAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGC 2197

RESULT 15  
US-09-514-340-1  
Sequence 1, Application US/09514340  
Patent No. 6361987  
GENERAL INFORMATION:  
APPLICANT: Akira ARISAWA et al.  
TITLE OF INVENTION: GENE ENCODING A PROTEIN HAVING ASYMMETRIC  
HYDROLASE ACTIVITY FOR 4-SUBSTITUTED 1,4-DIHYDROXYRINDINE DERIVATIVES  
ITS EXPRESSION PRODUCT  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.  
STREET: 2033 K Street, N.W., #800  
CITY: Washington  
STATE: D.C.  
COUNTRY: U.S.A.  
ZIP: 20006  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: Wordperfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/514.340  
FILING DATE: 28-Feb-2000  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 09/000.016  
FILING DATE: January 30, 1998

ATTORNEY/AGENT INFORMATION:  
NAME: Lee Cheng  
REGISTRATION NUMBER: 40,949  
REFERENCE/DOCKET NUMBER: <Unknown>  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-721-8200  
TELEFAX: 202-721-8250  
TELEX: <Unknown>  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2809 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: genomic DNA  
ORIGINAL SOURCE:  
ORGANISM: Streptomyces antibioticus  
STRAIN: <Unknown>  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 338...2539  
IDENTIFICATION METHOD: E  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 2540...2809  
IDENTIFICATION METHOD: P  
SEQUENCE DESCRIPTION: SEQ ID NO: 1:  
US-09-514-340-1  
Alignment Scores:  
Pred. No.: 1,09e-23 Length: 2809  
Score: 327.50 Matches: 136  
Percent Similarity: 40.83% Conservative: 51  
Best Local Similarity: 29.69% Mismatches: 160  
Query Match: 14.54% Indels: 111  
DB: 4 Gaps: 23  
US-09-985-689A-7 (1-433) x US-09-514-340-1 (1-2809)  
QY 18 GlyLeuTyrGlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArg 37  
Db 998 GGCTACGACGCGCAAGGCGGTGAAGATCGCGTCTCGACACCGGTCTCGACACGAGC--- 1054  
QY 38 AsnAspSerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGly 57  
Db 1055 -----CATCCGACCTCAAGGCGCGGTGACCGCTCCAGACATTCACC 1099  
QY 58 ArgThrAsnAsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeu 77  
Db 1100 GCCCGCGCGCGCGCGCGCAAGGTGGCGCCAGCCACCCACCGCTCGATCGCGCGC 1159  
QY 78 GlyAsnAlaThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuVal 92  
Db 1160 GGCACGGCGCGCCAGTCCAAAGGCAAGTACAGGGGTGCGCACCCGCGCGGATCTC 1219  
QY 93 PheGlnSerIleMet---AspSerGly---GlyGlyLeuGlyGlyLeuProAlaAsnLeu 110  
Db 1220 AACGCGCAAGGTCTTCGACGACTCCGGTTTCGGGACGACTCCCGCATCTCCGCGCGCATG 1279  
QY 111 GlnThrLeuPheSerGlnAlaTyrSerAlaGlyAlaArgIleHisThrAsnSerTrpGly 130  
Db 1280 CAGTGGCGCGCGCGCAG-----GGCGCGCGCGTCTCAACATGACCTGGGG 1327  
QY 131 Ala-----ProValAsnGlyAlaTyrThrThrAspSerArgAsn 143  
Db 1328 GGCATGGACACACCGGAGACCGCCCGCTGGAGCGCGC----- 1366  
QY 144 ValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAlaGlyAsnGluGly 163  
Db 1367 CTCGACAGCTGTCGCGCGGAGAGGGCGTCTCTGCCCATCGCGCGCGCACGAGGCG 1426  
QY 164 ProGlySerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrValGlyAla 183  
Db -----

Db 1427 CCGAGTCG-----ATCGGTTCGCGCGGAGCGGCGCGCCCTCACCGTCGCGCC 1480  
Qy 184 ThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsnHisValAlaGln 203  
Db 1481 -----GTGACGACCAAGCAAGCTCGCCGAC 1507  
Qy 204 PheSerSerArgGlyPro---ThrArgAspGlyArgIleLysProAspValMetAlaPro 222  
Db 1508 TTCTCTCCACCGCGCCCGCTCGGCGGACGCGGCCATCAAGCGGAGCTCACCGCTCCC 1567  
Qy 223 GlyThrTyrIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrpAlaAsn 242  
Db 1568 GCGTGGACATACGCGCGCTCGCGGAGGCAACGACATCGCGCAGGAGTCTCGTGAG 1627  
Qy 243 HisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIleValAlaGly 262  
Db 1628 GGACCGCGCGCTACATGACCATCTCCGGCACGTCGATGGCGACCCCGCACGTCGCGGGC 1687  
Qy 263 AsnValAlaGlnLeuArgGluHisPheValIleAsnArgGlyValThrProLysProSer 282  
Db 1688 CGCGCGCGCTCTCTGAAGCAGCAG-----CACCCCGAC 1720  
Qy 283 LeuLeuLysAlaAlaLeu-----IleAlaGlyAlaAlaAspValGlyLeuGlyPhe 299  
Db 1721 TGGACCTCCCGCGGAGCTGAAGCGCGCTCACCGCTCCACCAAGCGGCAAGTACACC 1780  
Qy 300 ProAsnGlyAsnGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsnValAlaPhe 319  
Db 1781 CCG--TTCGAGCAGGGTTCGGCGCGGATCCAGGCCGACAAAGCGCTCCAGCAGACCGTG 1837  
Qy 320 ValAsnGluThrSerProLeuSerThrSer----- 329  
Db 1838 ATC-----GCCGACCGGTCTCGGTGAGCTTCGGCGTCCAGCGTGGCGCGCACCGAC 1891  
Qy 330 -----GlnLysAlaThrTyrSerPheThrAlaGlnAlaGlyLysProLeu 344  
Db 1892 GACGAGCGGTCCACCAAGCAGCTGACCTACCGCAACCTCGGCACCCAGGACGTCACGCTG 1951  
Qy 345 LysIleSerLeuValTrpSerAsp-----AlaPro----- 354  
Db 1952 AAGCTGACGTGACCGCGCACCGCCCAAGGGCAAGCGCGCCCGCGGGCTTCTTCACG 2011  
Qy 355 ---GlySerThrThrAlaSerLeuThrIleuValAsnAspLeuAspLeuValIleThrAla 373  
Db 2012 CTGGCGCCACCCAG-----GTGACCGTC 2035  
Qy 374 ProAsnGlyThrLysTyrValGlyAsnAspPheThrAlaProTyrAspAsnAsnTrpAsp 393  
Db 2036 CCGCGCGCGCGCAGC---GCCTCCGTGACATGACCGCC-----GACCCCGGCTCGGC 2086  
Qy 394 GlyArg---AsnAsnValGluAsnValPheIleAsnAlaProGlnSerGlyThrTyrThr 412  
Db 2087 GGCACGGTGGAGCGCGGTACTCGGCGTACGTGTCGCC---ACGGCGGGGGGCGACAGC 2143  
Qy 413 Val-----GluValGlnAlaTyrAsnValProVal 422  
Db 2144 GTCCGACGCGCGCGGTGCGAGCGGAGGTGCGTACGTGACGAGCTGACCGTC 2197

Search completed: April 4, 2004, 12:04:03  
Job time : 92.4512 secs